Development and Validation of Assays used to Evaluate STAT3 Inhibitors

Chamberlain, Christopher Daniel

Awarding institution:
King's College London

Development and Validation of Assays used to Evaluate STAT3 Inhibitors

Christopher D Chamberlain

A thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy



Institute of Pharmaceutical Sciences

King's College London

2015

Supervisors:

Professor David E. Thurston

Doctor Khondaker Rahman

Abstract

Transcription factors are important control proteins in cells that bind to their cognate DNA sequences in the promoter regions of genes, either up-regulating or down-regulating protein expression. In many cancer types, transcription factors are up-regulated and promote the expression of genes important in survival and metastasis. For this reason, transcription factors are good targets for novel anticancer agents.

The STAT family of transcription factors (seven are now acknowledged) recognize and bind to a ~10 base pair sequence of DNA in the promoter region of a number of genes, enhancing the expression of oncogenic proteins such as Survivin, Cyclin D1, Bcl-2 and VEGF. There are currently no small-molecule STAT3 inhibitors in clinical use, so there is a need for the development of assays that can be used to screen molecules to identify lead compounds. The main focus of this project has been to develop an *in vitro* homogenous time resolved FRET (HTRF) assay that can be used in low-, medium- and high-throughput modes for the discovery of novel inhibitors.

The project started with the cloning, production and purification of recombinant STAT3β_{TC}, which is a homodimeric protein. This was challenging and time-consuming as initial solubility and stability issues were encountered. However, experimental conditions were eventually established that allowed useful quantities (i.e.10 mg batches) of purified and stable protein to be obtained. As part of the optimization process, the STAT3β_{TC} was re-cloned into a HIS-Tag vector which facilitated purification using affinity (Ni^{2+}) chromatography along with size exclusion chromatography to produce pure monomeric STAT3β_{TC}. This could be dimerised to provide pure STAT3β_{TC} homodimer.

The pure protein was used to develop a HTRF assay by first labelling the STAT3β_{TC} with Europium. Next, the cognate DNA recognition sequence in the form of an 18-mer duplex oligonucleotide was biotinylated and joined to the second fluorophore label (D2) via a streptavidin linkage. The strength of the FRET signal between these two components could then be used to measure the interaction between them. As part of a multi-well system, this could then be used to screen for small molecules capable of disrupting the protein/DNA complex. The assay was validated using unphosphorylated

STAT3 that does not form the biologically-relevant homodimer, and non-biotinylated DNA, which would not form the active FRET pair.

Further validation of the assay was carried out using known STAT3 inhibitors such as the peptidomimetics PYLKTK and YLPQTV, and the small-molecule inhibitors STA-21 and Stattic. It was then used to screen a 40-membered library of novel SH2-targeted molecules produced in-house, in which it successfully identified six “hit” molecules with low micro molar activity. These were further evaluated by establishing IC_{50} values in a number of cell lines including MDA-MB231, HELA, A4 and NCI-H1975. These studies revealed a correlation between the FRET assay results and the cytotoxicity of the molecules in the STAT3-dependent cell lines. The molecules were also studied in cellular experiments to establish their effect on STAT3-regulated genes such as Cyclin D1 and Survivin, in which a correlation was also observed. As a result, these molecules are now in further development.

Finally, the assay has been modified for high-throughput use in a 384-well system, and will be used for robotic screening in the future.

Table of Contents

TABLE of FIGURES

Table list

Acknowledgements

I am extremely grateful to a number of people without whom this work would not have been possible. Firstly, I would like to thank all the members (past and present) of the Cancer Research UK Protein-Protein interaction Drug discovery group led by Professor David Thurston at The School of Pharmacy and subsequently at Kings College London.

I would like to express my thanks to my Supervisors Dr Giovanna Zinzalla, Dr Khondaker Miraz Rahman and Professor David Thurston. Your guidance and support have been crucial.

From my days at SOP, I would like to thank Mrs Samantha Essex, Dr Emma Sharp, Dr Colin James, Dr Edwin Nkansah, Dr Giovanni Melillo and Dr Andrew Wilderspin.

From my days at KCL, I would like to thank Mr Jonathan Palmer, Mrs Kazi Rahman, Dr Paul Jackson, Ms Julia Mantaj, and Dr Thomas Cornell, who kept me on track.

On a special note, I would like to thank my wonderful family, siblings and parents. Paula and Keith, Chris and Rosy; this would have simply not been possible without your support. Holly and Calvin, you remind me what is important in life. Emily, you continue to amaze me. I love you all.

"When we long for life without difficulties, remind us that oaks grow strong in contrary winds and diamonds are made under pressure."

Peter Marshall

Abbreviations

ASO	Antisense oligonucleotide
AUC	Analytical ultracentrifugation
Bcl-XL	B-cell lymphoma-extra large
BP	Base pairs
BSA	Bovine serum albumin
CAD	C-terminal transcriptional activation domain
CNTF	Ciliary neurotrophic factor
CSF	Colony stimulating factor
CV	Column volume
DBD	DNA binding domain
DDW	Double distilled water
DMAP	4-Dimethylaminopyridine
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
cDNA	Complementary DNA
DTT	Dithiothreitol
EDC	1-Ethyl-3-(3-dimethylaminopropyl) carbodiimide
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
FP	Fluorescent polarisation
FRET	Forster resonance energy transfer
GAS	Gamma activated sequence
GH	Growth hormone
HNSCC	Head/neck squamous cell carcinoma
HPV	Human papilloma virus
HTRF	Homogenous time-resolved FRET

IC_{50}	Half maximal inhibitory concentration
IFN	Interferon
IL-6	Interleukin 6
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Iso-thermal calorimetry
JAK	Janus kinase
Ka	Association constant
LB broth	Luria-Bertani broth
LIF	Leukaemia inhibitory factor
MAPK	Mitogen-activated protein kinase
Mcl-1	Induced myeloid leukaemia cell differentiation protein Mcl-1
MMP	Matrix metalloproteinase
mRNA	Messenger ribonucleic acid
NAD	*N*-terminal transcriptional activation domain
dNTP	Deoxyribonucleotide triphosphate
OD	Optical density
OSM	Oncostatin M
PBS	Phosphate buffered saline
PDGF	Platelet derived growth factor
PDPN	Podoplanin
PIAS	Protein inhibitor of activated STAT
PTP	Protein tyrosine phosphatase
PTPRD	Protein tyrosine phosphatase receptor type D
PTPRT	Protein tyrosine phosphatase receptor type T
RCF	Relative centrifugal force
RPM	Revolutions per minute
SDS	Sodium dodecyl sulphate
SOCS	Suppressor of cytokine signalling

SH2	Src homology 2 domain
SPR	Surface plasmon resonance
STAT	Signal transducer and activator of transcription
TAD	Transactivation domain
VEGF	Vascular endothelial growth factor

1 General Introduction

1.1 Cancer

Almost one out of every two people born today will at some point in their lives be diagnosed with cancer (American Cancer Society, 2013). It is testament to the way modern medicine has reduced the threat posed from external threats to health that cancer in its many forms affects so many lives. In learning how to treat and prevent cancer, we are increasing our understanding of many branches of biology, from genetics and evolutionary biology to cell and developmental biology. It sometimes seems that the more questions we answer the more complex the intracellular environment appears and it is this complex system that has formed over billions of years that we must understand in order to come up with specific treatments for very specific errors in function.

Cancer is an umbrella term for over 200 different types of disease, named after the type of cell or organ from which it originates. Our bodies' cells must multiply and divide to replace old and damaged cells in what is normally a tightly controlled and regulated process. Cancer occurs when this process goes awry, with either cells not dying when they should or new cells forming when they should not. This accumulation of cells can have local and systemic effects, especially when they invade or metastasise to other tissues of the body.

The initiation and development of the cancerous phenotype involves an interaction, damage and/or change to the genomic DNA. These changes have been linked to many factors such as tobacco smoking, alcohol consumption, exposure to excessive sunlight or radiation, poor diet and exposure to carcinogenic chemicals or viruses (Human papilloma virus). The disease state develops further when cells are unable to regulate their division, proliferation and homeostasis. Cancerous phenotypes are the result of the accumulation of mutations to genes; these mutations favour selection for cells with aggressive phenotypes and in a time dependent manner lead to invasive malignancy (1). The progression of the disease state is

driven by the activity of two oncogene categories; tumour suppressor genes (recessive/loss of function) that regulate cell growth and proto-oncogenes (dominant /gain of function), which function as accelerators to activate the cell cycle (2). These oncogenes are active in all cancers and organ tumour subtypes.

A cancer cell exhibits phenotypes that include; loss of differentiation signal, unregulated cell division, loss of stop signal for proliferation, invasion of tissues and metastasis, sustained angiogenesis and evasion of apoptosis. These alterations allow tumour cells to stay viable, adapt and stay one step ahead of the bodies anticancer defence mechanisms (3).

The three main traditional approaches to treatment are surgery, radiation therapy and chemotherapy. Surgery and radiotherapy are currently the most widely and effective used local treatments, although ultrasonic and targeted radio wave destruction are becoming more routinely used. When the disease becomes metastatic however, chemotherapy is often the therapy of choice.

1.1.1 Current Approaches to Cancer Chemotherapy

Traditionally cytotoxic drugs have targeted either the cellular DNA or the mechanisms of cell proliferation. Alkylating agents intercalate between DNA bases covalently, preventing cell division by stopping the separation of DNA strands. Cytotoxic drugs may chemically change DNA structure (e.g. Platinum compounds) or substitute bases for inactive structures (e.g. 5-fluorouracil, gemcitabine). Drugs may deplete the supply of bases required for DNA and RNA synthesis (Methotrexate. 6-mercaptopurine, 6-thioguanine). Some drugs react directly with DNA-associated proteins such as topoisomerase I (e.g. topotecan, irinotecan) or topoisomerase II (e.g. etoposide). They may also affect microtubule formation during mitosis (e.g. taxanes, vinca alkaloids). Cisplatin works by intra- and inter-strand cross-linking of guanine residues; this causes inactivation of mismatch repair genes (4). All these approaches do little to distinguish between malignant and normally dividing cells, and,

therefore, cells with high turnover are highly impacted alongside the cancerous cells (e.g. bone marrow, hair follicle and gastrointestinal mucosa) and the search for ever increasingly specific drugs continues.

In the post-genomic era, pharmaceuticals are increasingly designed for a specific target as the understanding of cancer biology has increased (5). These novel targets are often focused on the deregulated control of cancer cell growth and division, and these signal transduction inhibitors are thought to have fewer side effects than traditional cytotoxics. The consequence of this less aggressive approach, however is that these agents may be primarily cytostatic and therefore used as a combination therapy where early disease diagnosis is essential.

Some forms of cancer are very hormone dependent (breast, prostate and ovarian cancers) so drugs that block the hormone from interacting with the cancer cell receptors are known as hormone therapies. Tamoxifen is commonly used in breast cancer and acts by blocking cell oestrogen receptors. Similarly, luteinising hormone blockers (e.g. Goserelin) and anti-androgens (e.g. bicalutamide) are used in prostate cancer by antagonizing testosterone receptors.

Various biological agents are used where more conventional drug compounds are ineffective. Monoclonal antibodies (MABs) work in various targeted ways,

Trastuzumab (Herceptin) binds to human epidermal growth factor receptor 2 (HER2) inhibiting the cancer cells ability to grow.

Bevacizumab is an anti-angiogenesis MAB that inhibits vascular endothelial growth factor (VEGF). Inhibition of overexpressed VEGF limits the blood and hence oxygen supply available to cancer tumours, retarding growth and subsequent metastasis.

1.1.2 Transcription Factors

Transcription factors associated with over activity in cancer cells are attractive drug targets as the products of their transcription are required for metastasis and cancer cell growth. An understanding of the eukaryotic transcriptional machinery is required before an approach at its inhibition can be attempted. Genes consist of three distinct regions, the promoter region, the enhancer region and the coding region itself.

Promoter sequences are DNA sequences that define where transcription of a gene begins. Promoter sequences are located at the 5' end of the transcription initiation site and indicate which DNA strand will be transcribed; this strand is known as the sense strand. Many eukaryotic genes have a conserved promoter sequence called the TATA box (6). This is a type of promoter sequence located 25 to 35 base pairs upstream of the transcription initiation site. Its conserved DNA sequence of TATAAA led to its naming, and transcription factors bind to the TATA sequence and initiate the formation of the RNA polymerase transcription complex, promoting transcription. A response element is a promoter or enhancer sequence closely associated with the gene it regulates. It is, therefore, known as a cis-acting element. Steroids or hormones released due to cellular stresses bind to response elements causing a gene expression response; p53 also forms response element complexes as part of its regulatory role as an activator protein (7). Enhancer regions act similarly to promoter regions but may be located either upstream or downstream of the gene it regulates. Also, an enhancer does not need to be located near the transcription initiation site to affect transcription; some have been found located in several hundred thousand base pairs upstream or downstream of the start site. Enhancers are bound by activator proteins. These activator proteins interact with the mediator complex, which recruits polymerase II and the general transcription factors, which then begin transcribing the genes.

Transcription factors may also be described as either general or gene specific transcription factors. General transcription factors, along with RNA polymerase II, assemble and bind on

the promoter regions of genes whereas gene specific transcription factors bind to the enhancer regions and contact RNA polymerase II via a large protein complex known as mediator. Mediator is essential to transcription of RNA polymerase II promoters. The gene specific transcription factors form a protein complex with coactivators which are responsible for acetylating the lysine residues of histones and, therefore, alleviating a physical barrier to transcription (8).

Many signalling pathways (~16) have been implicated in the control of cell division and apoptosis; these pathways are usually highly conserved and are often activated during embryogenesis and fetogenesis. Extracellular ligands (Interleukin-6, cytokines, growth factors) interact with receptor associated Janus kinases (JAKs) and Src family kinases which, through phosphorylation cascades, transmit the signal into the cell. The phosphorylated cytoplasmic proteins are then transported into the cell nucleus where they up or down regulate gene expression by acting as part of a transcription factor complex (9).These signal transduction steps are mediated at many points by the Protein-Protein interactions of transcription factor complexes. Octamer transcription factor (Oct-1) has been found to regulate the expression of the genes of the JAK-STAT signaling pathway (IFNAR2, STAT1, STAT2, and STAT4). Overexpression of Oct-1 isoforms caused a decrease in the activity of these JAK-STAT genes in a lymphoma cell line (10).

The activation of latent transcriptional regulators, allows an efficient and direct route for cellular communication from the receptors on the cell surface to the regulation of gene expression. Transcription factor pathways can be classified as in Fig 1.1 (11).

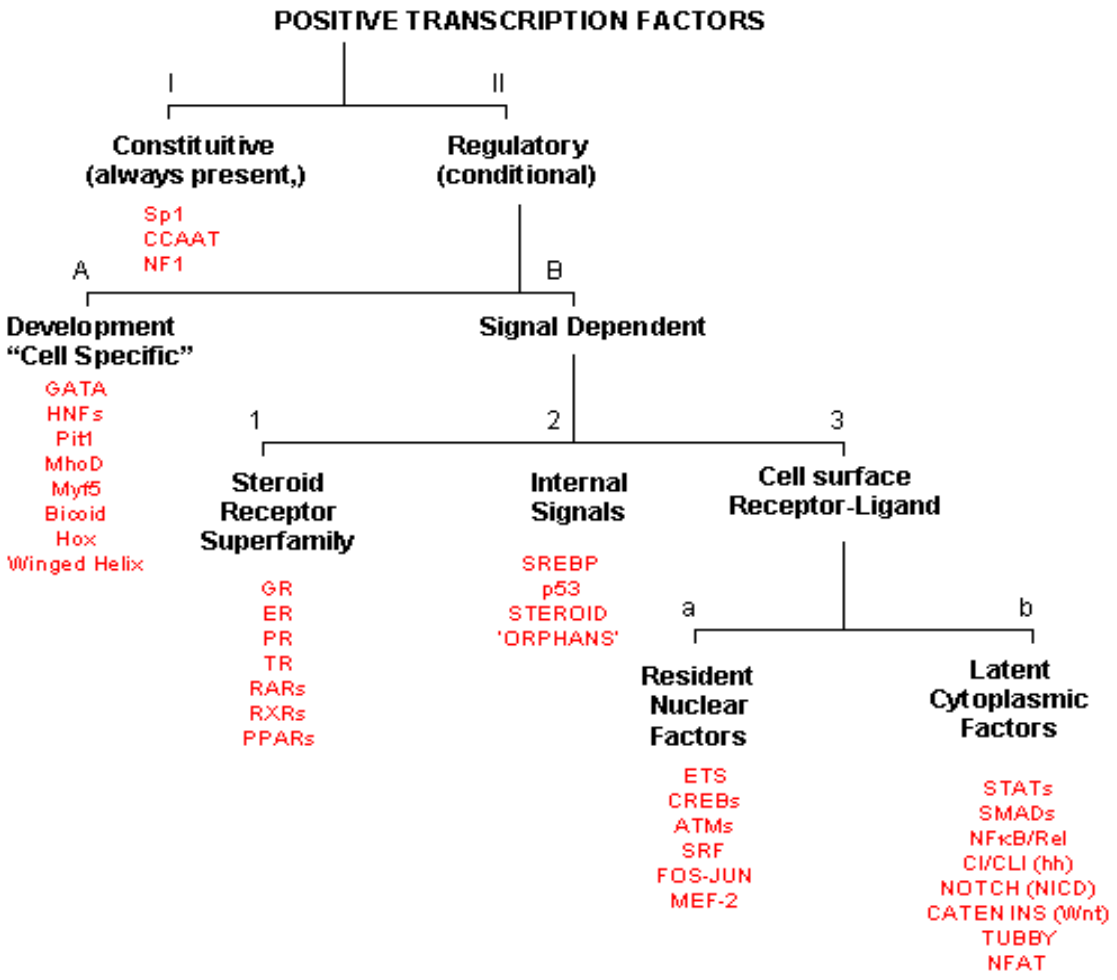


Figure 1.1 Schematic representation of the different classes of transcription factor (12)

Latent cytoplasmic transcription factors (LCTFs) are activated by a cell surface receptor-ligand interaction. Upon activation the transcription factor translocates into the cell nucleus where it interacts with other proteins of the transcriptosome in the regulation of gene expression. STATs are latent cytoplasmic transcription factors and are unique in their ability to respond to a cell receptor through to DNA binding (13).

1.1.3 Protein-Protein Interactions

Genetic studies have identified transcription factors as a rich source of potential targets for cancer therapeutics. While enzymes and receptors have long been modulated by drug-like molecules, the modulation of protein-protein interactions is more challenging. However, this

broad perception is not appropriate as we learn more of protein structure and with the aid of increasingly powerful modelling docking algorithms, these once featureless protein surfaces are being seen with greater resolution and detail.

Protein surfaces are characterized by numerous indentations and clefts which in the unbound state may associate with water molecules creating a unique environment of charged features commonly referred to as a “hotspot”. These hotspots are small specific areas of just a few amino acids which are crucial to the success of the entire interaction between two proteins; therefore, by targeting this relatively small but crucial area can be enough to prevent dimerization. Small drug-like molecules are capable of modulating a protein interaction of approximately 300-1000Å. The average size of a hot spot is 600Å (14) and so hot spots represent a valid therapeutic target.

1.1.4 Protein-Protein Interactions as a Drug Target

PPI specificity is important in organising the regulatory processes of cells and is commonly mediated by modular protein domains such as the Src homology domain (SH-2) (15) and Per Arnt Sim (PAS) domain. The inhibition of specific protein-protein interactions within pathways, especially those involved in signal transduction leading to carcinogenesis and tumour formation, are fast becoming attractive targets for novel anticancer therapeutics (16).

There are three main approaches to therapeutic targeting of a protein-protein interaction. Firstly, and most common, is to target a domain hot-spot, which is the approach taken in this study. However, allosteric regulation may be equally effective and may help to explain experimental results where inhibition does not agree with predicted modelling data. Thirdly, PPI stabilization rather than inhibition may also have the effect of down regulating transcription by reducing protein disassociation and subsequent re-activation.

Most PPI inhibitors share a structural similarity in the projection of lipophilic residues in a 3-D projection; this is thought to mimic the shape of either the α-helices or β-turns with which they are in competition (17).

There are a number of promising PPI inhibitors that are in development as chemotherapeutic agents showing that this area is steadily growing. For example, Navitoclax (ABT-263) (Abbott laboratories) is a dual inhibitor of anti-apoptosis proteins Bcl-2 and Bcl-X_L., the major negative regulators of apoptosis. The Bcl-2 proteins work by binding to two other groups of proteins, the executioners (Bax, Bak) that actually start the apoptotic pathway. Inhibiting this PPI induces apoptosis in the cancer cell. Navitoclax is also bioavailable orally and, in animal studies, it was found to induce apoptosis in senescent cells (17, (18, 19). ABT-737 (an ABT-263 mimetic) is in phase II clinical trials and has been found to be most effective when used in combination with a pharmacological up regulator of Noxa (a BH3-only pro-apoptotic BCL-2 family protein) such as Vorinostat (an HDAC inhibitor) (20).

Nutlin-3 is a specific small-molecule inhibitor of MDM2, which blocks the protein-protein binding of MDM2 with p53, activating the anticancer activity of p53 (21). Nutlin-3 also binds to Bcl-X, an anti-apoptotic protein. Forming a Bcl-X/Nutlin-3 complex in modelling studies and spectroscopic analysis shows Nutlin 3 to be a good example of competitive targeted PPI inhibition (22).

1.1.5 Drug Parameter Challenges

Four main classes of PPI modulators are currently considered:

1. Monoclonal antibodies- despite being expensive and not orally bioavailable, antibodies are highly specific and represent the majority of PPI inhibitors currently in use in the clinic (23).
2. Peptides and peptidomimetics- poor stability and low oral bioavailability, although advances are being made into improving metabolic and pharmacokinetic properties

(24). Peptides are often used as molecular probes and antagonists for investigating PPI modulators.

3. Small molecule PPI modulators- drug like molecules that fall within Lipinski's rule of five.
4. Unconventional PPI modulators- These molecules do not conform to classic drug-like classification, but, through advances in drug delivery, are very much worth further investigation (25).

1.2 Signal Transducers and Activators of Transcription (STATs)

STATs (signal transducers and activators of transcription) have been found in many biological systems from unicellular *Dictyostelium* (slime moulds) to mammals and so are thought to have been highly conserved from early eukaryotes (26). JAK/STAT signalling appears to be an early adaptation to facilitate intercellular communication that has co-evolved with a myriad of cellular signalling events. This co-evolution has given rise to highly adapted, ligand-specific signalling pathways that control gene expression.

STATs are characterized as a group of proteins found in the cell cytoplasm that are activated by extracellular polypeptides interacting with cell membrane receptors in an interconnected series of pathways of PPIs (27) . The latent STATs are subsequently phosphorylated, which enables them to form homo- and hetero-dimers. The dimers then relocate to the nucleus where their effect on gene transcription and hence regulation is achieved. They do not require secondary messengers to convey a signal from the cell membrane to activate transcription and this makes them unique.

STATs were first identified by Darnell, Kerr and Stark in a study of their activation with interferon (21,(28), although many polypeptide activators have now been identified. As the name suggests, the family has dual functionality of transducing an extracellular signal into an intracellular one and then subsequently activating transcription.

The STAT protein is phosphorylated on a tyrosine residue, a key residue in the homodimerisation at the Src-Homology 2 (SH2) domain. The dimer then enters the nucleus where it binds to DNA promoter regions and regulates transcription of many genes; STATs are bound by an ability to convert an extracellular receptor signal into a genetic response.

Currently seven mammalian STATs have been identified: STAT1, STAT2, STAT3, STAT4, STAT5a, STAT5b and STAT6. In humans the genes are spread across three chromosomal clusters (Table 1.1).

	Gene loci	Activators	Cancers	Oncogenes
STAT1	Chromosome 2, band q12 to q33		Breast, head and neck, lung, brain	
STAT2	Chromosome 17, band q11-1 to q22			
STAT3	Chromosome 12, band q13 to q14-1		Breast, head and neck, prostrate, melanoma, thyroid, myeloma	*Src, eyk, ret, lck, $G_{\alpha}o$, Npm-alk*
STAT4	Chromosome 2, band q12 to q33	IL-12, IL-23		
STAT5a	Chromosome 12, band q13 to q14-1		Chronic myelogenous leukaemia, Acute myeloid leukaemia, acute lymphoblastic leukaemia, Erythroleukaemia	lck
STAT5b	Chromosome 12, band q13 to q14-1		Chronic myelogenous leukaemia, Acute myeloid leukaemia, acute lymphoblastic leukaemia, Erythroleukaemia	lck
STAT6	Chromosome 17, bands q11-1 to q22			

Table 1.1 STAT loci, activators and cancer implications, adapted from (29)

1.2.1 STAT Sequences and Domains

STAT molecules share distinct domains that link them in structure and function. These have been determined through comparing gene sequence and mutagenesis. STATs 1, 3, 4, 5A and 5B are between 750 and 795 amino acids long whereas STATs 2 and 6 are larger, at approximately 850 amino acids (30). The first 130 amino acids at the amino terminal of STATs are involved in multiple STAT cooperative binding of DNA consensus sequences where multiple copies exist in the genome; these sites are thought to be specific for the different STAT proteins and may exist to direct selective transcriptional activation (31, 32). However, the first 130 amino acids of the NH_2 terminus are not required for STAT dimerization or subsequent binding to a single consensus sequence so are often omitted in recombinant studies. The next ~190 amino acids (130 to 320) form a bundle of 4 α-helices known as the coiled coil domain, followed by ~150 amino acids (320 to 470) which make up an eight stranded β-barrel known as the DNA binding domain, but, although it confers specificity, it is not enough to bind DNA on its own (33). A ~110 amino acid (470 to 580) linker region follows, made up of 4 short α-helices, which lead onto the ~135 amino acid (585 to 720) SH2 domain important in STAT protein-protein interaction in dimerisation via phosphorylation at tyrosine 705. The C terminal domain is a transcriptional activation domain whose activity is enhanced by phosphorylation of serine 727 (Fig 1.2).

1.2.2 STAT3 Isoforms

Splicing variations are known to exist in all seven of the known STAT genes resulting in the expression of a number of additional proteins. These isoforms exist in different cell types and may be expressed in order to regulate STAT function. STAT3 has a beta (β) isoform that has a shortened TAD domain at the C-terminal of the gene, where 55 amino acids from the α form are replaced by a 7 amino acid sequence (34). The β isoform therefore lacks the phosphoseryl residue (S723), but retains the tyrosine (Y705) critical for homodimerisation in

the SH-2 domain (35). The phosphorylation of tyrosine 705 occurs constitutively, so the dimer is very stable and constitutively binds to the DNA consensus sequence, hence regulating gene expression (36). The gene set is different between the two isoforms; *in vivo*, the β isoform may exist as a regulator of the α-isoform (37). A study of green fluorescent protein-tagged Stat3α and Stat3β demonstrated that the two isoforms have very different intracellular activity, with Stat3β shown to be retained in the nucleus for longer and less nuclear mobile, especially following ligand stimulation. This makes STAT3β the preferred isoform drug study target as its exposure to DNA consensus sequences is greater (38).

1.2.3 STAT3β

Splicing variations are known to exist in all seven of the known STAT genes resulting in the expression of a number of additional proteins. These isoforms exist in different cell types and may be expressed in order to regulate STAT function (39).

STAT3β is a truncated isoform of STAT3α in that the last 55 C-terminus amino acids are replaced by a 7 residue string (34)34, (40).

STAT3β was first observed for its DNA-binding activity in IL-6 stimulated hepatocytes as an acute–phase response factor (41, 42). It was later discovered that this same protein is activated by the entire family of IL-6 like cytokines, which signal through gp130 and related receptors (43). Similar to other STATs, STAT3β has a conserved amino-terminus involved in tetramerisation (31), a DNA-binding domain.

As well as homodimerisation, STAT3β can form a heterodimer with STAT1 (44), although specificity is conferred to different cytokine activation profiles (45).

STAT3β has long been described as an oncogenic transcription factor that mediates oncogenic transformation in cultured cells and tumour formation in nude mice (46). STAT3β protects the cancerous cell and progresses the malignant phenotype by up-regulating key

proteins. Cell survival (anti-apoptotic) proteins such as Bcl-2 and Mcl-1 protect the cell from destruction. Bcl-2 promotes cell survival by inhibiting cytochrome C, which is an important signal in the apoptosis pathway. Cyclin D1/D2 and c-Myc regulate the cell-cycle and proliferating cell division (47). The cyclins regulate transition through the cell cycle stages through regulation of cyclin-dependent kinases and subsequent complex formation. Vascular endothelial growth factor (VEGF) induces angiogenesis ensuring the malignant cell maintains its supply of oxygen and nutrients.

1.2.3.1 STAT3β Structure

STAT3β's binding relationship to its consensus DNA sequence was first shown along with STAT1 in 1998 (48, 49). An adaptation of this structure is seen below in figure 1.2. This resolution at 2.25Å (PDB 1BG1) shows STAT3β homodimer bound to DNA covering residues aa127 to aa722, the truncated form used in this study. The elongated 4-helix coil domain (aa127-aa320) is shown in magenta. The coiled-coil domain at the amino terminus contains four antiparallel α-helices and is essential for Stat3 recruitment to the receptor and the subsequent tyrosine phosphorylation and tyrosine phosphorylation-dependent activities, such as dimer formation, nuclear translocation, and DNA binding (50). The DNA binding domain (DBD)(aa320-aa465) is shown in red interacting with the bound DNA (cyan); it forms the physical binding region with STAT3' target DNA consensus sequence and has been linked to binding to NF-kappaB p65 in order to inhibit NF-kappaB-mediated transactivation of the inducible nitric oxide synthase (iNOS) gene (51). The linker domain (aa465-aa585) is in gold and sits between the DBD and the SH2 domain with both being structurally compromised if mutations are introduced to the linker, affecting pSTAT3-dependent transcription (52). The SH2 domain in blue (aa585-aa722) incorporates the C terminal domain (C-TAD) in yellow (aa688-aa722), a major tyrosine phosphorylation site at Y705. Phosphorylation leads to dimerization of STAT3 via intermolecular pTyr-SH2 interactions. The STAT3 beta isoform (84 kDa) is missing from this domain (1-715 + 7 unique amino

acids resulting from frameshift) and is sometimes used as a dominant negative though there is also evidence that it regulates distinct genes as well.

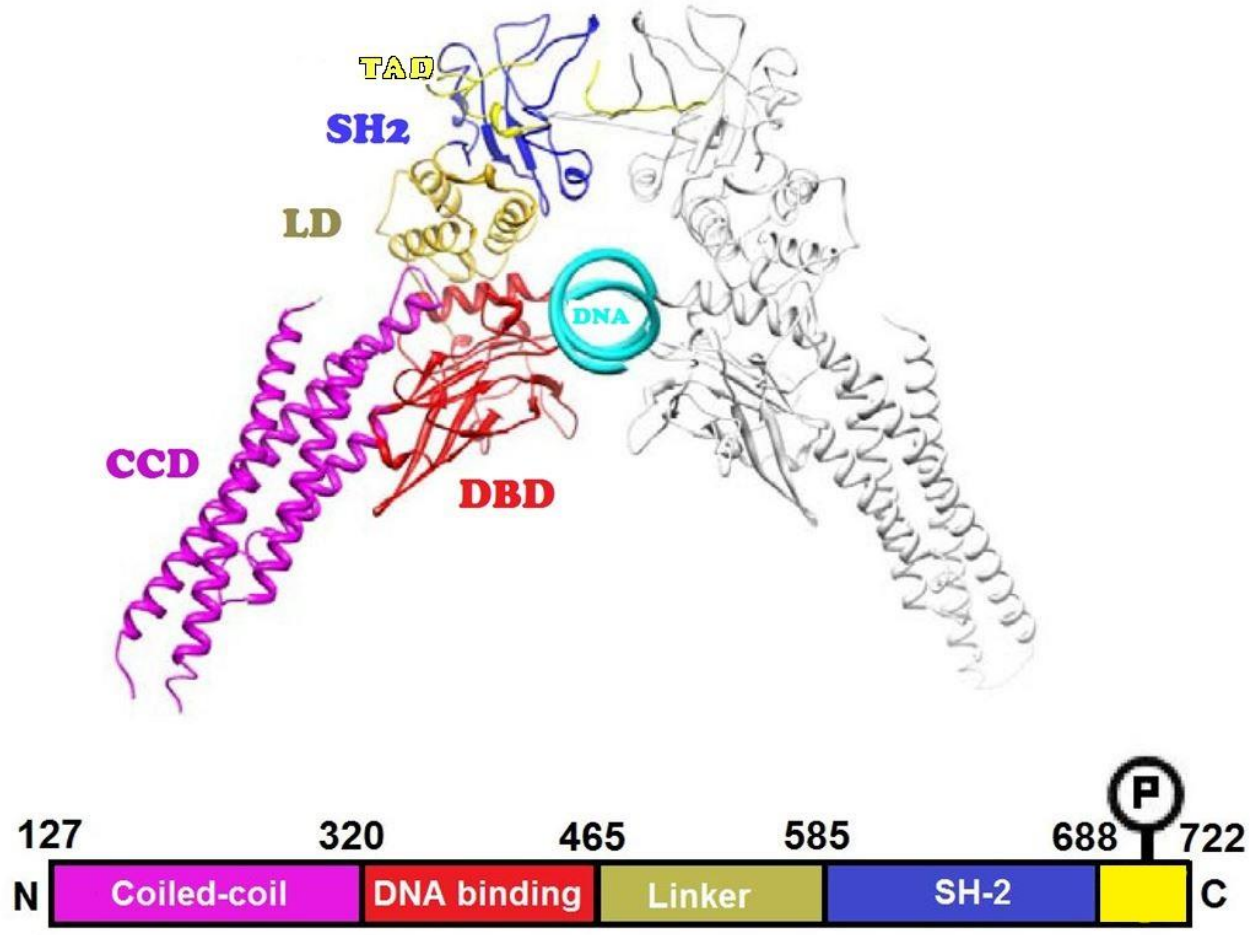


Figure 1.2 Model of the crystal structure of STAT3β homodimer bound to DNA and domain map showing region size {adapted from Muller (48)}. The *N*–terminal coiled-coil domain (CCD) is shown in magenta, the DNA binding domain (DBD) in red, the linker domain (LD) in gold, the SH2 domain (SH2) in blue and the C-terminal domain (TAD) in yellow. The double stranded DNA (*ds*M67) is shown in cyan. The structure was adapted from PDB/1BG1 (53).

1.2.3.2 STAT3 Physiological Functions in Normal Cells

STAT3 first gained recognition as an acute phase response factor that was activated by IL-6 (54), but now the number of activating compounds has grown to include other cytokines (e.g. IL-7, IL-10 and IL-20), and also growth factors, granulocyte colony stimulating factor and leptin. This physiological function of effecting cytokine and chemokine receptor signaling led to signal transducer and activator of transcription (STAT) naming.

In normal biology, several different mechanisms are responsible for STAT3 down regulation. A family of proteins named suppressors of cytokine signaling (SOCS) inhibit STAT3 by binding to upstream Janus kinases and inhibiting their phosphorylation of STAT3 (55, 56). Another family, called protein inhibitors of activated STAT (PIAS) (57), are ligases that act similarly to ubiquitin by specifically binding STAT3 and blocking its activity (58, 59).

STAT3 down regulation is also controlled by the enzymatic de-phosphorylation of Y705 through protein tyrosine phosphatases (PTPs). These enzymes can be either integrated into the cell membrane (PTPR) or are cytosolic (PTPN). In addition deacetylases (I.e. SirT1) play a part in STAT3 down regulation through the removal of K685's acetyl group (60).

STAT3's complex network of activation and deactivation regulation illustrates its importance in normal cell biology homeostasis and it has been ubiquitously present in a high proportion of the bodies' tissues.

STAT3' role in normal tissue is most understood in the immune and inflammation response (1, 61). It was the response to Il-6 and interferon (IFN) that initially led to STAT3's discovery, where the pro-inflammatory activity is reactionary and transient (self-regulated through SOCS3, itself up-regulated by STAT3). The anti-inflammatory response to Il-10 however results in a more sustained activation of STAT3 (62). High STAT3 activity within normal cells is also associated with wound healing, again through IL-6 up regulation at sites of cell damage. Indeed IL-6 knockout mice have a heal time three-times longer than the wild type (63).

STAT3 has been implicated as a downstream effector of hormone systems in the brain such as insulin and leptin, indicating a role of STAT3 in the homeostasis of metabolite and energy levels (64, 65).

Partly due to its ubiquity, STAT3 has been shown to play a role in embryogenesis; in fact, STAT3 knock out mice die early in embryogenesis, prior to gastrulation.(66) Mammary development, thymic function and autophagy are also highly dependent on STAT3 activity (66-69). This activity in normal tissue is tightly controlled and transient. Its role in cancer whereupon this regulation goes astray is where its interest as an anticancer drug target originates. STAT3 has been shown to inhibit the p53 gene transcription rate through binding to the p53 promoter in vitro and in vivo. p53 protein is a potent inhibitor of cell growth, arresting cell cycle progression at several points and inducing apoptosis of cells undergoing uncontrolled growth. (70)

1.2.4 STAT3 in Cancer

Most human cancer cell types have been shown to have overexpressed STAT3 protein levels (71). As a key communication transducer of signals from extracellular ligands through to specific genome expression, many mutations can result in abnormal STAT3 levels; here I summarize some of the main oncogenic targets and their effects.(72)

1.2.4.1 Genomic Deregulation of STAT3

The over activity of STAT3 in human cancers is an effect of the deactivation of tumour suppression proteins and activation of oncoproteins. Mutation of the STAT3 gene itself is very rare (1.08% of tumours of ~5000 tumours analysed by The Cancer Genome Atlas) (73). Abnormal regulation of the many pathways that regulate STAT3, causes the normally highly

regulated and transient STAT3 signaling to become unregulated, leading to angiogenesis, proliferation, invasion and survival.

In the disease state upstream kinases become mutated, causing constitutive activation of STAT3 (e.g. myeloproliferative neoplasms are often the result of the JAK2 mutation V617F) (74). Gene amplification of epidermal growth factor receptor (EGFR) occurs in glial tumours which in turn lead to greater activation of wild-type receptor kinases and subsequent lasting STAT3 activation (75-77).

Similarly epigenomic inactivation of STAT3 down regulating proteins in cancer cells consequently leads to over activity of STAT3. For instance protein tyrosine phosphatases receptor type D and T (PTPRD and PTPRT respectively) are often changed in cancer cells (through methylation of their promoter region) inhibiting their ability to regulate pSTAT3 through de-phosphorylation (78)

By targeting STAT3 directly a wide range of upstream abnormal changes to both activators and deactivators of STAT3 in cancer cells can be minimized.

1.2.4.2 STAT3 in Proliferation and Growth

Unregulated STAT3 activation leads to transcriptional up regulation of genes encoding Cyclin D1 and MYC. Cyclin D1 acts through a Cyclin-dependent kinase to facilitate the movement of the cell through the G1 stage of the cell cycle. In this way excess Cyclin D1 leads to perpetual cell grow and proliferation (79). The transcription factor Myc acts to promote growth in the cell. Myc is up regulated by STAT3, therefore causing premature progress through the cell cycle in many cancers (80).

1.2.4.3 STAT3 in Cell Survival and Apoptosis

STAT3 is implicated in the transcriptional over-expression of anti-apoptotic proteins such as Bcl-xL and Bcl-2 (81, 82). The Bcl-2 family regulates cytochrome c release in the mitochondria which is responsible for the initiation of apoptosis. The over-expression of the anti-apoptotic Bcl-2 protein in lymphocytes alone does not cause cancer. However, simultaneous over-expression of Bcl-2 and the proto-oncogene Myc may produce aggressive B-cell malignancies including lymphoma (83). The transcription of excessively high levels of Bcl-2 through STAT3 decreases the propensity of cells for apoptosis (84).

1.2.4.4 STAT3 in Malignancy and Invasion

STAT3 has been shown to cause an increase in metastasis to the brain in melanoma (85). This metastasis is initiated by matrix metalloproteases (MMP's); these endopeptidases degrade extracellular matrix proteins. Tumour metastasis is a multistep process involving the release of tumour cells from the primary tumour to secondaries at a distant organ or tissue (86). One of the first steps in metastasis is the degradation of the basement membrane releasing anchored cells from their clustered state. MMP-2 and MMP-9 genes are up-regulated by STAT3, and STAT3 knockdown in esophageal squamous carcinoma cells leads to MMP down-regulation, dysregulation of cell migration and decreased migration speed (87).

Transmembrane glycoprotein Podoplanin (PDPN) is similarly overexpressed through STAT3 over activity in squamous cell carcinoma cells. PDPN is believed to play a key role in the cancer cell invasiveness by mediating efficient extracellular matrix degradation (88). Increased cell migration and invasion due to increased cell density can be reversed by RNA knockdown of PDPN (89).

1.2.4.5 STAT3 in Angiogenesis

STAT3 works to promote angiogenesis in two distinct ways; firstly STAT3 causes up regulation of pro-angiogenic factors such as vascular endothelial growth factor (VEGF) and IL-6, which leads to an autocrine feedback loop (90). STAT3 also acts as a downstream effector of cytokine receptors such as leptin, IL-6 and VEGF. Inactivation of STAT3 *in vivo* enhanced the death of the cytokine-dependent sensory neurons of the nodose ganglion, demonstrating that STAT3 signalling plays a role in mediating the survival response of neurons to cytokines (91). Increased angiogenesis is crucial to tumour development, ensuring that energy and oxygen are in ready supply to the dividing cells (92).

1.2.4.6 STAT3 Activation and Regulation

More than 35 different polypeptide ligands, along with oxidative stress (93), have been identified to interact with cell surface receptors to activate the dormant intercellular STATs (94). In a simple flow path STATs are phosphorylated at a specific tyrosine residue on the Src-homology 2 (SH2) domain. A mutual SH2 phosphotyrosine interaction occurs between two STAT monomers. The homodimer complex is then able to pass into the nucleus and bind to promoter consensus sequences to effect transcription of target genes. A summary table of STAT3 activators and the kinases they activate is given in Table 1.2.

Ligand (family)	Phosphorylating enzyme
Gp 130	
IL-6	
IL-11	
Oncostatin M (OSM)	
Leukaemia inhibitory factor (LIF)	
Ciliary neurotropic factor (CNTF)	
yC family	
IL-2	**JaK1, JaK3**
IL-7	**JaK1, JaK3**
IL-9	**JaK1, JaK3**
IL-15	**JaK1, JaK3**
Interferon (IFN)	
IFN-α/β/ω	**Tyk2, JaK1**
IL-10	**Tyk2, JaK1**
IL-20	**?**
IL-22	**?**
Single chain family	
Growth hormone (GH)	**JaK2**
Receptor tyrosine kinases	
Epidermal growth factor (EGF)	**JaK1, JaK2**
Platelet derived growth factor(PDGF)	**JaK1, JaK2**
Colony stimulating factor (CSF-1)	**Tyk2, JaK1**
Hepatocyte growth factor(HGF)	**?**

Table 1.2 STAT3 activation ligands and associated enzymes {adapted from (95)}

1.2.5 JAK/STAT Signaling

The Janus kinase-signal transducer and activator of transcription (JAK-STAT) pathway represents a major mechanism used to transmit signals from extracellular receptors to the nucleus. The canonical pathway consists of JAK and STAT proteins activated by receptor ligation and inactivated by negative regulators, including SH2-containing protein tyrosine phosphatase (SHP) and suppressor of cytokine signaling (SOCS) proteins.(96)

The Janus kinase (JAK) signaling pathway starts outside the cell with the binding of the ligands given in Table 1.2 to its associated receptor. The cross membrane receptor subunits are bound together either as homodimers (erythropoietin and growth hormone) or heterodimers (interferon and interleukins). Upon extracellular ligand binding the receptor complex undergoes a conformational change that allows intracellular binding of two of the receptors associated JAKs. The close proximity of the JAKs allows *trans*-phosphorylation of the receptor endodomain at multiple tyrosine residues (97).

These receptor phosphotyrosyl motifs recruit cytoplasmic STAT3 via its SH-2 domain and once in association with the receptor/JAK complex, the active JAK phosphorylates the STAT3 (98). The phosphorylated STAT3 disassociates from the enzyme/receptor complex where it is free to homodimerise with a partner pSTAT3. Once dimerised, the complex diffuses to the nucleus where, in association with importin-α, the complex is transported across the nuclear pore by the Ran nuclear import pathway (99, 100).

Once transported into the nucleus, dimerised STAT3 binds to genomic regulatory sequences of DNA in order to either “switch on” (i.e. VEGF) or “turn off” (i.e. p53) the transcription of target genes. In this way the JAK/STAT signalling pathway provides a crucial mechanism of transducing an extracellular signal ligand into a transcriptional change (101).

1.2.6 Positive Regulation of STAT3

In addition to phosphorylation through the JAK pathway, STAT3 has been shown to be phosphorylated at Y705 through activation of non-receptor tyrosine kinases (v-Src and v-Abl) (102), hormones (insulin and angiotensin) and guanine nucleotide-binding proteins (G-protein receptors) (103, 104). Upon activation, dimers bind to DNA consensus sequences (TT $(N)_{4-6}$AA) and regulate gene transcription of specific target genes. A 2013 study of our group also showed that unphosphorylated STAT3 could bind to the DNA consensus sequence and, thereby, validating unphosphorylated STAT3 as a drug target in inhibiting downstream gene expression (105).

A reporter gene study (106) showed that phosphorylation of serine residue 727 by mitogen-activated protein kinase (MAPK) in a conserved C-terminal PMSP motif is essential for prolonged maximum STAT3 activation. This extra activity suggests that there is collaboration between different signalling cascades regarding STAT3 (107, 108). For example MAPK is activated by the Ras pathway (109, 110).

microRNAs (miRNAs) are small non-coding RNAs that bind to complementary sequences on target mRNAs, often silencing gene expression. miR-19a has been shown to regulate suppressor of cytokine signalling (SOCS1) expression, thus enhancing STAT signalling (111).

1.2.7 Negative Regulation of STAT3

The phosphorylation of JAKs and STAT3 are negatively regulated by protein tyrosine phosphatases (PTPs). A tyrosine phosphatase (SHP-1) contains two SH-2 domains and binds to phosphorylated JAKs and phosphorylated receptors to initiate de-phosphorylation of these activated signalling complexes. This upstream de-phosphorylation regulates the downstream phosphorylation of STAT3 (112).

A second group of negative STAT3 regulators are suppressors of cytokine signalling (SOCS). The SOCS family of proteins become associated with target proteins (JAKs and STATs) through their SH-2 domain, inhibiting JAKs and competing with STAT for phosphotyrosine binding sites. The bound STAT-SOCS complex is targeted for ubiquitination and the polyubiquitinated complex is in turn targeted by the proteasome for degradation (113). SOCS proteins form a negative feedback loop in this way, as they are themselves overexpressed through STAT3 target genes and act to turn off the pathway (114). For example SOCS2 inhibits STAT5 phosphorylation in response to growth hormone signalling. Hepatocytes derived from SOCS2 knock-out mice have prolonged STAT5a and STAT5b phosphorylation in response to growth hormone. The mechanism by which SOCS2 regulates STAT5 activation is still not completely understood. However, it appears to involve the competitive binding of SOCS2 to STAT5 and binding sites on the growth hormone receptor (115).

Protein inhibitors of activated STAT (PIAS') act primarily in the nucleus inhibiting STAT3 homodimers from binding to DNA consensus sequences. The mechanism for this is not clear, but it is thought that sumoylation (small ubiquitin-like modification) is involved (59). Low protein and mRNA expression levels of PIAS3 have been found in gastric cancer tissues compared with surrounding healthy tissue, indicating an important role in STAT regulation. (116)

Cleavage of STAT proteins by proteases is a form of proteolytic processing and results in the generation of C-terminally truncated proteins, called STAT gamma, which are missing the transactivation domain (CTAD) and behave as functional dominant-negative proteins, playing a role in regulation during times of STAT3 overexpression. STAT gamma isoforms have been identified for STAT3, STAT5a, STAT5b and STAT6 in different cellular contexts and biological processes (117).

1.2.8 The phosphorylated STAT journey

The Src-homology 2 domain (SH-2, amino acids 575-680) was first recognised in 1996 (118) and is fundamentally conserved through all STAT proteins, but has also been observed in many other functionally different proteins (i.e. enzymes, adaptors, regulators and docking proteins) (119). The SH-2 phosphotyrosyl domain binding is a polar interaction that occurs between residues Lys591, Arg609, Ser611 and Ser613 of one STAT3 monomer and Y705 of another activated STAT3 monomer (120). This binding interaction facilitates a conformational change in the dimer that allows nuclear translocation and subsequent DNA binding to occur (121). Activated STAT3 homodimers, once transported into the nucleus, bind to a consensus DNA sequence which is a member of the interferon-gamma activated sequence (GAS) family of transcription enhancers (122, 123). Biochemical analysis has shown that STAT3 has a strong binding affinity for the sequence TTCN2-4GAA, with the optimum being the 9 base pair palindrome (TTCCNGGAA) (124).

1.2.9 STAT3 Targeting Ligands

STAT3 has long been considered a desirable and “druggable” target, and various STAT3 inhibitors have been described with varying success and specificity (125). A STAT3 inhibitor must work by preventing STAT3 dimerisation, either by inhibiting STAT3 translocation into the nucleus or inhibiting STAT3 DNA binding (126). The inhibition needs to be specific enough that other STAT family members are unaffected and should not target the JAK phosphorylation event, as this is implicated in many other phosphorylation events. A clear distinction needs to be made between a STAT3 inhibitor and STAT3 pathway inhibition. STAT3 pathway inhibitors are compounds that cause reduced activation of STAT3 by indirect means and I will touch briefly on the advances in this field. This study however is primarily interested in STAT3 inhibitors, compounds that inhibit the STAT3 mechanism of action through direct protein binding.

1.2.9.1 Inhibitors of STAT3-STAT3 Dimerization

STAT3 monomers have been shown to dimerise *in vitro* independently of associated receptors (127) making possible the study of complex inhibitors. Inhibitors of the STAT3-STAT3 homodimerisation event generally target the SH2 domain of STAT3 and specifically the region of amino acids around Y705 as the phosphorylation of this tyrosine residue is of great importance to the dimerisation event. The 3D structure model of the STAT3β homodimer suggests that dimerization of STAT3β occurs primarily at the SH2 domain. The SH2 domains are hinged to each other by a looped segment (from Ala-702 to Phe-716) within each monomer. The phosphorylated tyrosine residue (Y705) crucial to the biological action of STAT3 forms an attraction to this loop segment and binds along with several adjacent amino acid residues (Leu-706, Thr-708, and Phe-710) to an accommodating pocket on the other SH2 domain. Specific binding by a small molecule to this residue on either or both of the STAT3 monomers would block further docking through steric hindrance and consequently block the dimerization of STAT3.

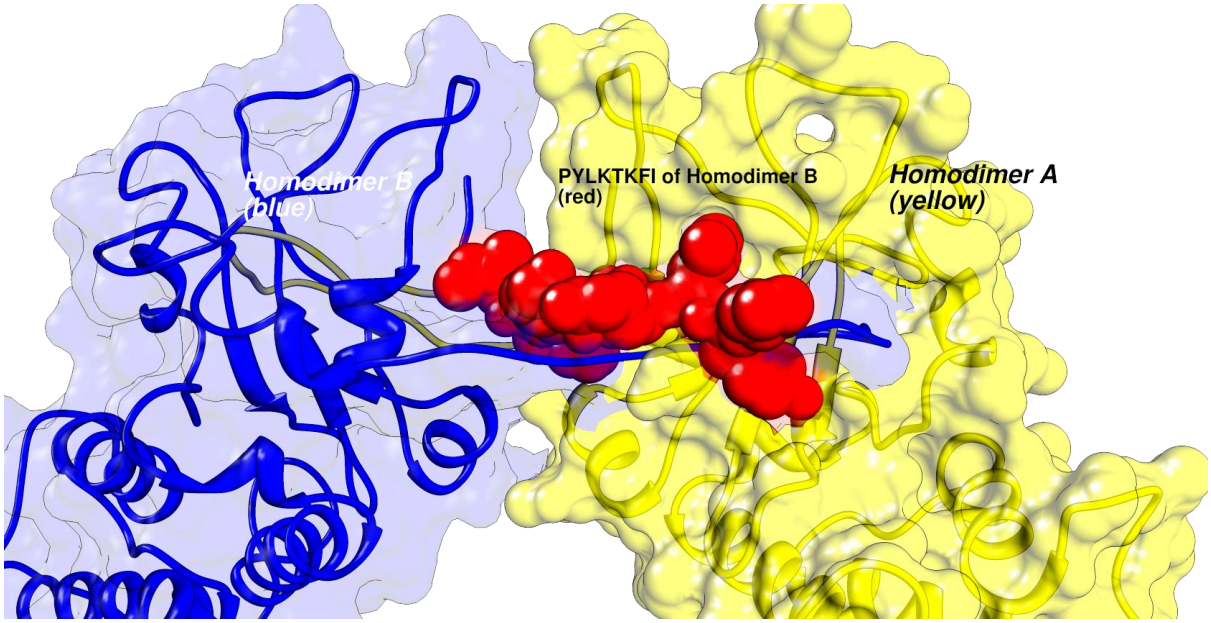


Figure 1.3 Molecular model of PYLKTKFI phosphopeptide (red) docked between STAT3 monomers (blue and yellow) (courtesy of Dr Paul Jackson)

This principle was first displayed by Turkson *et al* in 2001 where the presence of PY*LKTK (Y* being phosphotyrosine) disrupted STAT3 activity *in vitro.* The minimum active sequence was mapped to be the tripeptide XY*L (X being any substituent amino acid) and when associated with a membrane *trans* locating sequence (AAVLLPVLLAAP) to aid transportation into the cell; this was found to inhibit selectively constitutive and ligand-induced STAT3 activation *in vivo* (128).

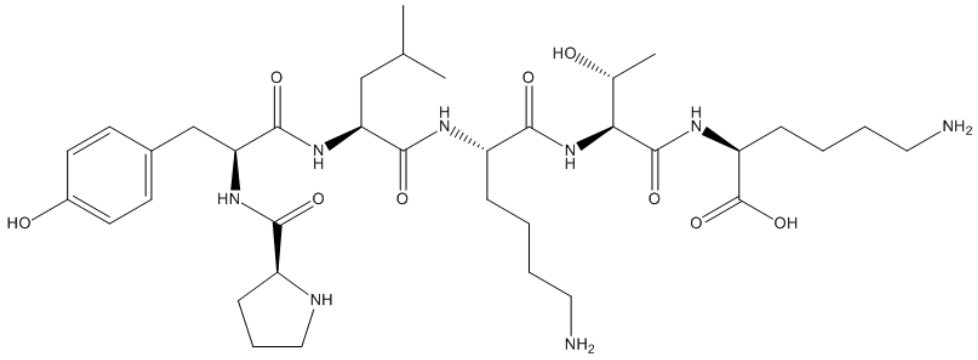


Figure 1.4. Structure of PYLKTK peptide

A number of tripeptide derivatives based on PY*L and AY*L (where Y* represents phosphotyrosine) have been developed in order to be more drug like and condensed than the full peptides. ISS- 610 and ISS-840 have a 4-cyanobenzoate substitution at the Y-1 position and were found to be over 5 times more potent than the native tripeptide. ISS-610 caused apoptosis and inhibited cell growth in Src transformed mouse fibroblasts and human lung carcinoma cells (128-130).

A peptidomimetic, Ac-Tyr (PO3H2)-Leu-Pro-Gln-Thr-Val-NH2, was found to bind optimally to the primarily hydrophobic SH2 region utilising lipophilic amino acid groups at the *N*-terminus providing a backbone structure to optimise. Further optimisation resulted in hydrocinnamoyl-Tyr (PO3H2)-Leu-cis-3,4-methanoPro-Gln-NHBn, with increased affinity along the molecule and improved IC_{50} of 125 nM versus 290 nM for the original peptide (131-133).

Peptide inhibitors are metabolically unstable, have poor permeability into cells and are therefore considered at the moment to be unsuitable for clinical use (134). The challenge subsequently has been to find small molecules that have a greater bioavailability and efficacy and a number of non-peptide small molecules have been reported as inhibiting STAT3 by directly binding to the SH-2 domain (135).

STATTIC (Figure 1.7) was discovered in 2006 by a group in Germany (131, 136). STATTIC was unique as it inhibited STAT3 SH2 domain function regardless of whether the STAT3 was phosphorylated or not. They screened a 17,000 small molecule library for an ability to compete with a fluorescein-labelled, phosphotyrosine-containing peptide specific to the STAT3 SH2 domain. One hundred and forty four compounds were found to inhibit phosphopeptide-protein interactions by >60%. A secondary inhibition screen looking at IL-6 driven STAT3 transport into the nucleus and inhibition of phosphorylated STAT3-DNA binding identified STATTIC as the front running compound. It is selective for STAT3 over STAT1 and STAT5. However, STATTIC was found to be susceptible to nucleophilic attack as the compound lost its inhibitory action in the presence of DTT. Inhibition also increased over time suggesting that alkylation of STAT3 was occurring. A cysteine residue on the opposite side of the protein from the phosphopeptide binding face is thought to be the modified residue and, if so, STATTIC is not a direct competitor of phosphopeptide binding; its inhibition may instead be due to an altered conformation of the SH2 domain.

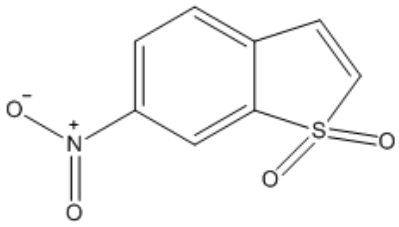
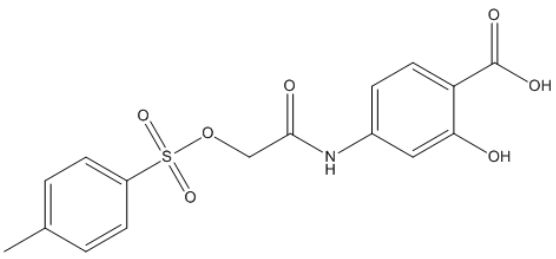


Figure 1.5 Structure of STATTIC

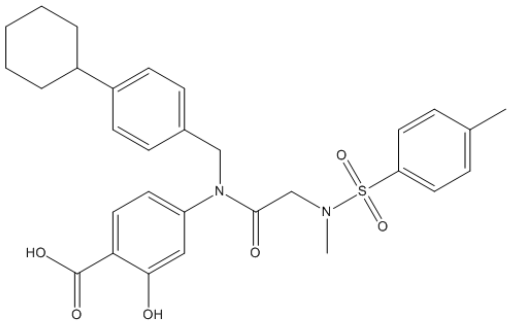
The screening of a library of molecules (Wyeth's) led to the discovery of a SH2 domain inhibitor that had an IC_{50} of 106 μM in a myeloma cell line. This catechol (1, 2-dihydroxybenzene) moiety was found not to be permeable to cells, but has proved useful to the adaption of drug like structures (137).

S3I-201 (also known as NSC74859) was also identified as a STAT3 dimerisation inhibitor through SH2 domain association in a NCI library screen, with an IC_{50} of 60–110 μM. S3I-201 was shown to inhibit STAT3 DNA-binding, alter STAT3' transcriptional profile, caused death in STAT3 dependent tumour cells, and slowed growth in human breast cancer xenografts (138).

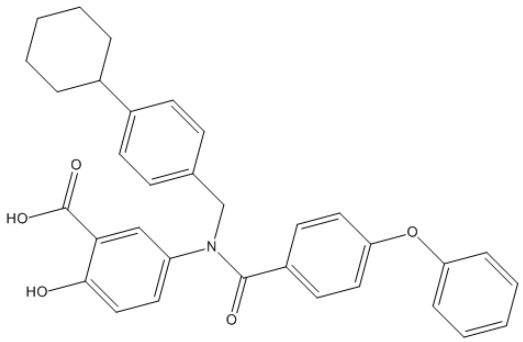
Greater potencies were seen in several S31-201's derivatives, including S3I-201.1066, and S3I-1757, with IC_{50} values of 35 μM, and 13.5 μM respectively, in constitutively active STAT3 cancer cells (16, 139).



A)



B)

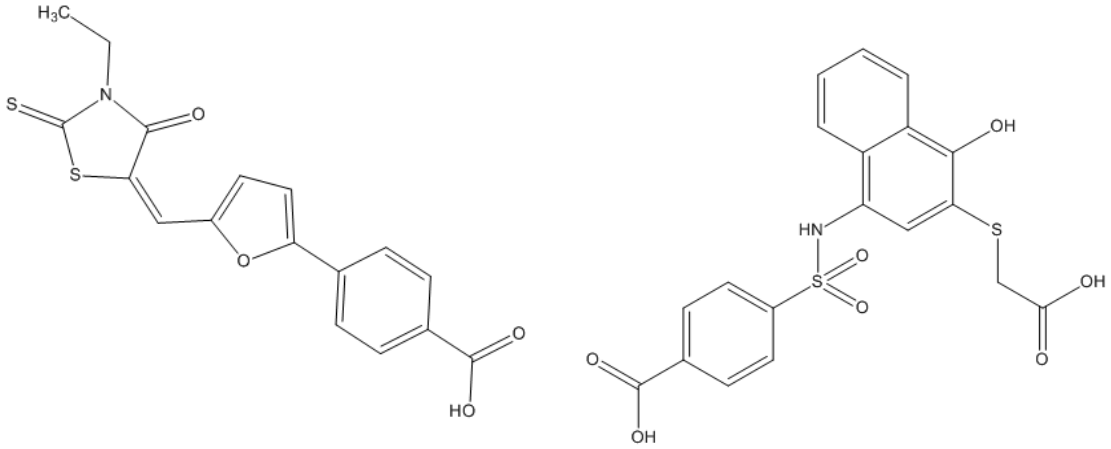


C)

Figure 1.6 Structures of A) S31-201, B) S31-201-1066 and C) S31-1757

Of these derivatives S3I-201.1066 was shown to have good potency and cause senescence in human breast and mouse non-small cell lung xenografts (140) (Figure 1.6).

A modelling library screen identified Cpd30 (4-(5- ((3-ethyl-4-oxo-2-thioxo-1, 3-thiazolidin-5-ylidene) methyl)-2-furyl) benzoic acid) and Cpd188 (4-((3-((carboxymethyl) thio)-4-hydroxy-1-naphthyl) amino) sulphonyl) benzoic acid). Cpd 30 blocks STAT3 nuclear translocation and induces apoptosis in STAT3-dependent breast cancer cells (141). Cpd-188 decreased tumour growth in resistant breast cancer xenograft models when used in combination with docetaxel (142).



A) B)

Figure 1.7 Structures of A) Cpd-30 and B) Cpd-188

A natural compound, cryptotanshinone (from *Salvia miltiorrhiza)* is thought to inhibit STAT3 functions through SH2 domain binding, although this requires further confirmation. The compound down regulated STAT3 regulated genes in prostate cancer cells (143).

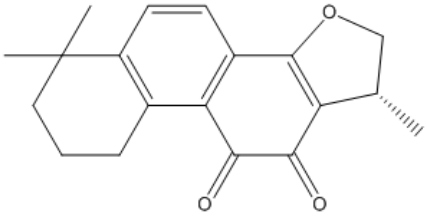


Figure 1.8 Structure of cryptotanshinone

Honokiol (figure 1.11) is a naturally-occurring compound found in extracts of magnolia bark. It has an acceptable bioavailability profile and has been shown to cross the blood-brain barrier. In an orthotopic mouse model, honokiol significantly decreased lung tumour growth compared with the vehicle control group. In a brain metastasis model, honokiol inhibited metastasis of lung cancer cells to the brain to approximately one third of that observed in control mice. Honokiol's mechanism of action is inhibition of STAT3 phosphorylation, and knockdown of STAT3 was seen to nullify the antimetastatic effects of honokiol. (144)

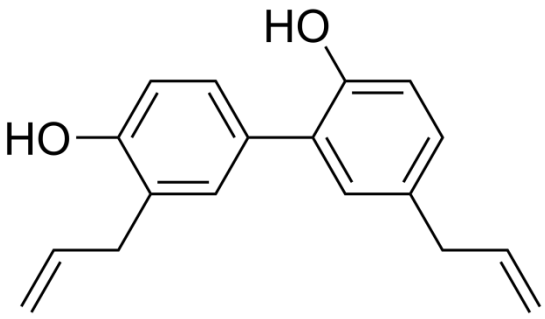


Figure 1.9 Structure of honokiol

Crispene E, a clerodane-type diterpene, was shown to inhibit STAT3 dimerization in a cell-free fluorescent polarisation assay and was found to have significant toxicity against STAT3-dependent MDA-MB 231 breast cancer cell line and selectively inhibited the expression of

STAT3 and STAT3 target genes cyclin D1 and bcl-2. Molecular docking studies suggest the molecule inhibits STAT3 by interacting with its SH2 domain (145).

1.2.9.2 Inhibitors of STAT3 DNA Binding

STA-21 (Figure 1.10) (146) was first published in 2005 by a group in Michigan who virtually screened over 400,000 database compounds, the top 100 of which were evaluated using a STAT3 dependent luciferase reporter assay. STA-21 inhibits STAT3 DNA binding, STAT3 dimerisation and reduces the survival of breast carcinoma cells with constitutive STAT3 signalling. STA-21 has been shown to decrease the progression of rheumatoid arthritis in mice through a decrease in T-cells. TNF-α and IL-6 production were decreased, along with affecting mRNA expression levels of key factors such as NF-κB, JAK1, STAT3 and p65 (147, 148).

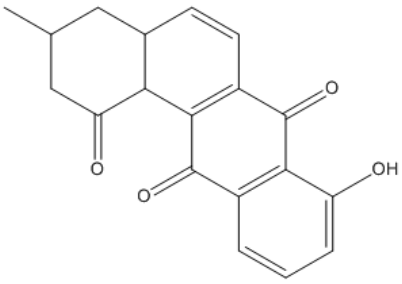


Figure 1.10 Structure of STA-21

A STA-21 structural analogue called LLL-3 (Figure 1.11) is significantly more membrane permeable, more cytotoxic *in vitro*, and decreased viability of intracranial cancer cells in a glioblastoma animal model (149). The acetyl group of LLL-3 was then replaced with sulphonamide to give LLL-12

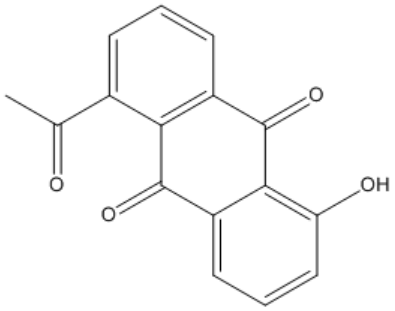


Figure 1.11 Structure of LLL-3

Oligodeoxynucleotide (ODN) decoys are one approach to inhibiting STAT3 DNA binding that is showing great promise in several clinical trials. ODNs work by competing with endogenous sequence equivalents for transcription factor association and, therefore, prevent gene expression. A 15-mer STAT3 decoy sequence (5'-CATTTCCCGTTAATC-3') replicates the *sis*-inducible element of the FOS promoter. Its toxicity was evaluated in a non-human primate and encouragingly no organ toxicity was seen during a two-week observation with intramuscular injection (150). A clinical trial of this ODN decoy (ClinicalTrials.gov identifier: NCT00696176) saw a single dose injected intra-tumourally in Head/Neck Squamous Cell Carcinoma (HNSCC) patients resulting in a reasonable safety profile and suppression of STAT3 target genes (151).

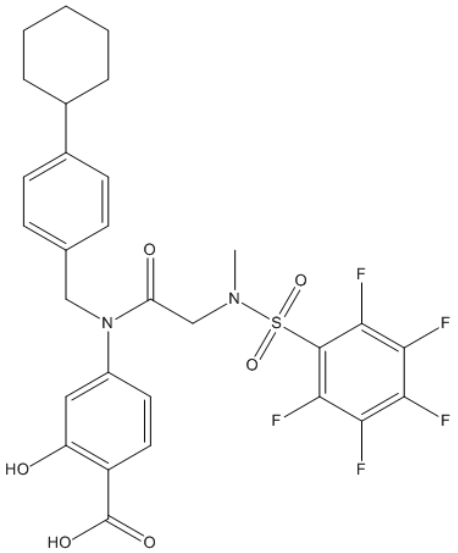


Figure 1.5 Structure of BP-1-102

BP-1-102 binds STAT3 with an affinity (KD) of 504 nM, blocks STAT3-phospho-tyrosine (pTyr) peptide interactions and STAT3-DNA binding at 6.8 μM (IC50). (139) A cyclic modification that is resistant to serum nucleases (5'-CATTTCCCGTAAATC-3') decreased tumour growth and down regulated STAT3 target genes in mouse glioblastoma xenografts and suppressed HNSCC and bladder cancer cell viability (151, 152). A hairpin version containing 2 STAT3 binding sites is able to select specificity for STAT3 over STAT1 and was active in a colon cancer cell line (SW480) (153).

A novel probe (inS3-54) showed selective inhibition of STAT3 binding to DNA without affecting the activation and dimerization of STAT3 through interaction with STAT3' DNA binding domain. InS3-54 also inhibits expression of STAT3 downstream target genes and STAT3 binding to chromatin *in situ* (154).

1.2.9.3 Inhibitors of STAT3 mRNA Translation

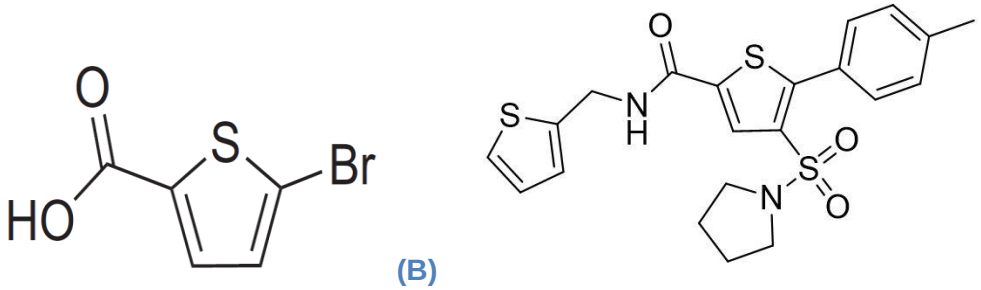
Antisense oligonucleotides (ASOs) work by targeting deviant STAT3 signalling. Single stranded complimentary versions of the overexpressed mRNA of downstream STAT3 gene targets inhibit their expression and then induce a blockade in the transfer of genetic information from DNA to protein (155). An early clinical study showed that ASOs specific to

STAT3 inhibited STAT3 expression and caused vascular endothelial growth factor (VEGF) down regulation in melanoma and breast carcinoma models (90). ASOs also suppressed malignancy in hepatocellular carcinoma and prostate cancer cells (156, 157).

ISIS pharmaceuticals (ISIS 481464) designed an ASO sequence targeting human STAT3 mRNA which was well tolerated in cynomolgus monkeys at 10 mg/Kg per week and over a 6 week period decreased STAT3 protein levels by up to 90% (158).

1.2.10 Thurston/Rahman Library Construction

A group developed scaffold, RH-06 (Fig 1.13) showed a moderate ability to disrupt dimerisation of the STAT3 protein by interacting with the SH2 domain (159). RH-06 inhibited STAT3 dimerisation in a fluorescent polarisation (FP) based primary protein-protein interaction assay with an absolute inhibition of 25.4%, compared to 41% observed for the natural ligand pYLKTKF. As RH-06 showed moderate STAT3 dimerisation inhibition by interacting with the SH2 domain, it was decided to use this molecule as a chemical scaffold and develop more potent and selective inhibitors of STAT3 dimerisation. Our group's aim to develop novel inhibitors of the STAT3 signalling pathway led to a library of arylsulphonamidyl thiophene amides being prepared after promising *in silico* observations. The library was based on a commercially available thiophene scaffold (Figure 1.13).



(A) (B)

Figure 1.6 Initial thiophene scaffold (A) and RH-06 intermediate (B).

This compound first underwent chlorosulfonation, followed by the formation of sulfonyl amide derivatives. R1 groups were attached with an amide bond by 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) coupling or 4-dimethylaminopyridine (DMAP) esterification to give initial intermediates. A final Suzuki reaction with various boronic acids gave the range of final ligands (Fig 1.12).

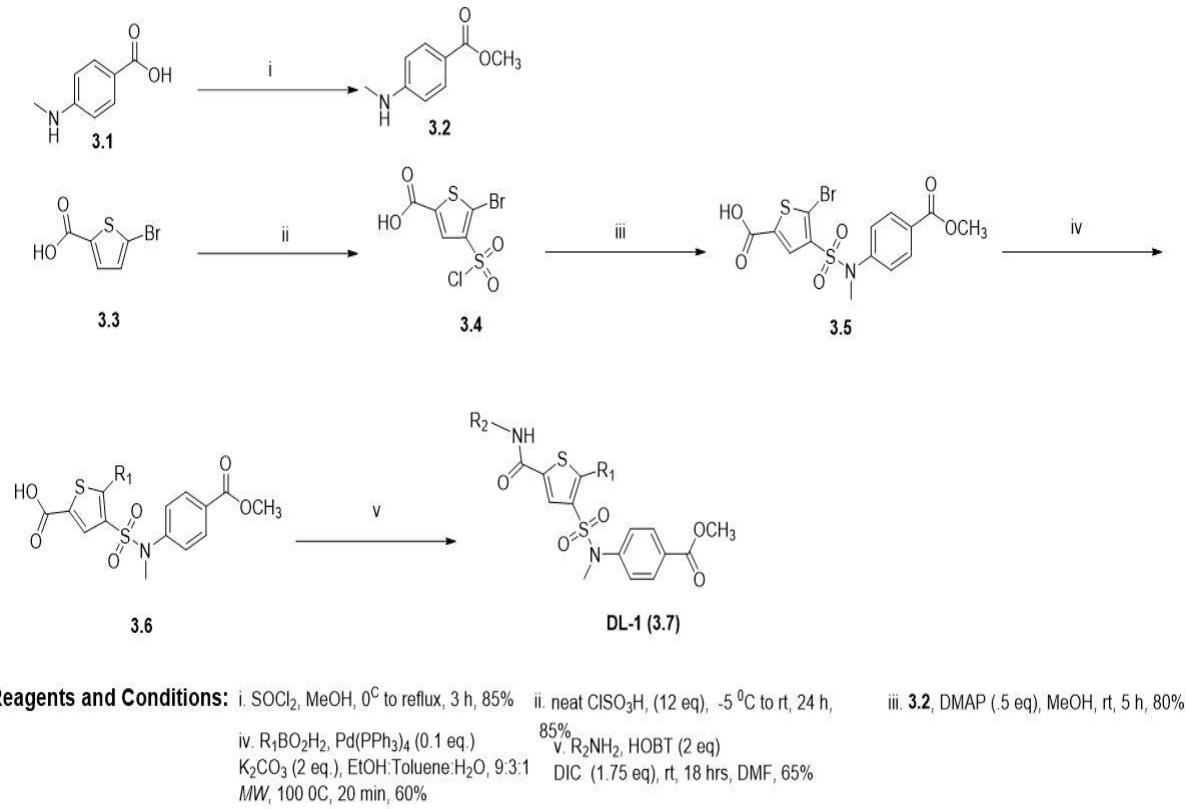


Figure 1.7 Synthetic scheme of compound formation (Courtesy of Mrs Kazi Rahman)

The target compounds synthesized were first confirmed by mass through LC-MS. Determination of structure was confirmed by NMR, FTIR and HRMS. Additionally, two separate HPLC gradient solvents were used to determine the purity of each compound.

1.3 Techniques

1.3.1 Cell Free PPI Characterisation

The study of protein-protein interactions starting with analytical techniques such as co-immunoprecipitation and affinity chromatography has been progressing at a phenomenal rate. Techniques such as analytical ultracentrifugation (AUC), light scattering, surface plasmon resonance (SPR) and iso-thermal calorimetry (ITC) create a broad spectrum of approaches. This study however is focussed on fluorescence spectroscopy techniques (160). Fluorescent screening technologies now make up the majority of high-throughput assays due to both the sensitivity and the high adaptability of the fluorescent signal.

1.3.2 Homogeneous Time-Resolved Fluorescence (HTRF)

Forster resonance energy transfer (FRET) is the observation that an excited donor fluorophore is able to transfer this excitation state, via a dipolar interaction to a suitable acceptor fluorophore (161). The distance between the two fluorophores must not be too great (~10Å – 50Å) and there must be a spectral overlap between the donor emission and the acceptor absorption. FRET has a greater chance of occurring when the excited state of the donor is long lasting and this is particularly utilised in time-resolved FRET. The fluorescence emission must occur in an area of the spectrum as different as possible from that produced by proteins; for example, a near infrared emission is well placed to avoid proteins medium-intrinsic fluorescence. When the fluorescent probes are placed on different biological subunits i.e. transcription factor complex and consensus DNA, the technique is particularly robust when studying molecular interactions (162).

A homogeneous immunoassay that utilized the FRET effect was first performed in 1976 (163). HTRF is a homogenous method which has the advantages of FRET whilst eliminating background fluorescence and quenching effects by introducing a time delay (50-150µs) between the fluorophore excitation and the emission measurements. Background fluorescence represents the baseline non-specific signal due to unbound fluorescence probe or sample auto-fluorescence. This baseline signal is subtracted from every sample test.

A transfer of energy occurs between a fluorescent donor (e.g. Europium stabilised in a cryptate cage) and an acceptor fluorophore such as d2 or XL665. In this HTRF assay, the donor and acceptor fluorophores are conjugated to biomolecules (anti-His antibody and streptavidin) in order to study molecular interactions. The two fluorophores used in this study were a europium fused 6HIS antibody and a streptavidin fused d2.

Europium cryptate (Eu3+cryptate) acts as the donor fluorophore in this FRET pairing; this complex consists of a macrocycle within which the Eu3+ ion is embedded. The cryptate cage allows both energy collection and transfer to the Eu3+ ion, which releases the energy with a specific fluorescent wavelength. Cryptates are formed by the inclusion of a cation into a tridimensional cage. The cage acts as a light collecting device and relays the energy to the core lanthanide ion. These properties of the macrocycle favour such a tight association with the lanthanide ion that this interaction becomes virtually unbreakable and leads to an exceptionally inert complex. . In comparison, chelates are not stable in acidic media and prone to exchange their rare earth ions with ions present in the media, like Mn2+. Cryptates are, therefore, less likely than chelates to dissociate the europium ion from its cage, This type of structure supports long-lived fluorescence, making Eu3+cryptate very suitable for time resolved studies.

The first generation of acceptor developed for HTRF was XL665, a phycobiliprotein pigment purified from red algae. XL665 is a large heterohexameric edifice of 105 kDa, cross-linked after isolation for better stability and preservation of its photophysical properties in HTRF

assays. This acceptor fulfils the compatibility criteria mentioned above. Its excitation spectrum overlaps that of Eu3+ cryptate emission, therefore allowing the donor to excite the XL665, and its maximum emission at 665 nm spans a region where Eu3+ cryptate does not emit or only does so weakly. This study uses the second generation of acceptor (d2), characterized by organic structures 100 times smaller, but displaying a series of photophysical properties very close to those of XL665. The comparison of d2 with XL665 was achieved by screening 14,700 compounds in an assay for quantifying a phosphorylated peptide. The correlation between the two systems was extremely close, and validated the integration of d2 in different HTRF assays. As a much smaller entity, d2 limits the steric hindrance problems which may occur in XL665-based systems.

The development of fusion proteins has enabled HTRF assays to use purified recombinant proteins by utilising the sequence tags on the protein. These tags are often present at the - terminal ends to aid in the identification and purification of the recombinant protein.

6HIS tag (six consecutive histidine amino acids) was used in this study for both purification (with a metal chelating solid phase) and assay development. Mouse monoclonal antibody HIS-1 is an IgG2a raised against polyhistidine tagged fusion protein. It is specific to synthetic polyhistidine or polyhistidine-tagged fusion proteins.

Streptavidin is a tetrameric protein (~60kDa) isolated from *Streptomyces avidinii* that binds strongly and with high affinity to biotin, a 244 Da vitamin found in the blood. It is an extremely strong non-covalent bond (Ka $=10^{-15}$M). The association is quick to form and the complex is virtually unbreakable under normal biological conditions. As biotin is a fairly small molecule, it can be conjugated to biomolecules without affecting their activity (164).

1.3.3 Cell Culture

This study used three cancer cell lines, HeLa, MDA-MB-231 and A4.

HeLa cells derive from human cervical cancer cells. They were originally cultured in 1951 and were the original human cancer cell line. HeLa cells are extremely fast growing given the correct conditions and nutrients, due to their metastatic nature. HeLa cells are cancerous due to infection with human papilloma virus 18 (HPV18), which is an initiator of cervical cancer. These cells are effectively immortal (no Hayflick limit) as the uncontrolled production of telomerase ensures that the chromosomal telomeres do not shorten after each division.

MDA-MB-231 is a breast cancer cell line derived from breast carcinoma which originated from the pleural effusion of a 51-year-old woman in 1973. Sixty per cent of breast cancer carcinomas are associated with constitutively active STAT3 (165). MDA-MB-231 is described as being triple negative (for ER-, PR- and no HER2 overexpression). Triple negative breast cancer (TNBC) is a subtype of breast cancer that is usually associated with poor outcome and lack of benefit from targeted therapy.

A4 is a STAT-null human DLD1 colon carcinoma cell line generated by homologous recombination (166) that is used in this study as a positive control for any cytotoxic effects seen in the MDA-MB-231 STAT3 dependent breast cancer cells.

1.3.4 MTT Assay

In this study an MTT assay was used to measure the cytotoxicity of the compound library. MTT assay has been used in STAT3 inhibition studies to support observations seen with *in vitro* (167, 168). A 2010 study found that the use of MTT as a measure of cell proliferation and viability can underestimate the anti-proliferative effect of compounds when compared to methods which measure metabolic activity through quantifying ATP (i.e.CellTiter-Glo) and cellular DNA content (i.e CyQUANT) MTT is useful in this context of a large scale screen, but once compounds are narrowed down any positive result seen may actually be more accurately seen with an ATP or DNA measuring screen (169)

An MTT assay is used to estimate proportionally the number of viable cells present. Viable cells are able to reduce the yellow tetrazole MTT to the reduced purple formazan by mitochondrial oxidoreductase enzymes (Figure 1.15).

The insoluble formazan is solubilised in dimethyl sulfoxide (DMSO) to produce a coloured solution that can be colourimetrically assessed by measuring the wavelength at 570 nm on a spectrophotometer.

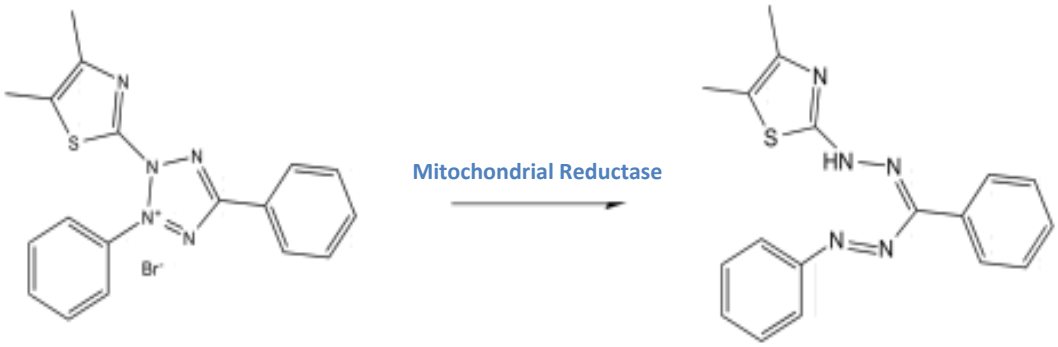


Figure 1.8 Reduction of MTT to formazan

2 Materials and Methods

2.1 Biological Materials

2.1.1 Laboratory Reagents

All laboratory chemicals and organic solvents were of research/analytical grade and were purchased from reputable suppliers (Table 2.3).

Supplying company	Reagent/Chemicals/Kits
Bio-Rad (Hertfordshire, UK)	Bio-Rad Bradford protein assay reagent
Fisher Scientific (Leicestershire UK)	Luria Bertani Broth (LB Broth) Agar powder Casamino acids Organic solvents (Ethanol, Methanol, Propan-2-ol) Bovine Serum Albumin (BSA)
Genron (Berkshire, UK).	NuSep coomassie brilliant blue G-250 dye Instant brilliant blue stain
Melford laboratories (Suffolk UK)	TRIS Base [Tris(hydroxymethyl)aminomethane] HEPES [*N*-(2-Hydroxyethyl)piperazine *N*'-(2-ethanesulphonic acid] Ampicillin sodium salt Dithiothreitol (DTT) Isopropyl-β-D-galactopyranoside (IPTG)
National Diagnostics (Yorkshire, UK)	Protogel (30% [w/v] acrylamide: 0.8% [w/v] bis-acrylamide)

New England Biolabs (Hertfordshire, UK).	1kb DNA marker Pre-stained Broad range protein marker
OXOID (Hampshire, UK)	Phosphate buffered saline (PBS) tablets
Promega (Southampton, UK).	PCR nucleotide mix (dATP, dCTP, dGTP and dTTP)
QIAGEN (West Sussex, UK).	Miniprep kit Maxiprep kit PCR clean-up kit Gel extraction kit
Sigma-Aldrich (Poole, UK)	*N,N,N′,N′*-tetramethylethylenediamine (TEMED) Tween-20 Ammonium persulphate Dimethyl sulphoxide (DMSO) Boric acid Ethylenediaminetetraacetic acid (EDTA) 3,3′-Diaminobenzidine (DAB) tablets SIGMA*FAST*™ Protease Inhibitor Tablets Antifoam Protease inhibitor cocktail for Histidine tagged proteins Phenylmethanesulphonyl fluoride (PMSF) Kanamycin Tetracycline Chloramphenicol Ethidium bromide Igepal Triton X100

	Agarose powder
	3-Indoleacrylic acid (IAA)
	Imidazole
VWR International Ltd	Sodium Chloride
(Dorset, UK)	Sodium Phosphate
	Ammonium Acetate
	Ammonium sulphate
	Potassium Chloride
	Sodium Hydroxide
	D(+) Glucose
	Glycerol
	Glycine
	Magnesium Chloride

Table 2.3 Laboratory reagents suppliers

2.1.2 Bacterial Expression Plasmids

The pET-32a(+)-STAT3β_{TC} expression vector containing the nucleotides sequence coding for the 127–722 amino acid residues of murine STAT3β_{TC} (identical to the human STAT3 on the protein level) was a gift from Professor C.W. Müller (EMBL Grenoble, France). The sub-cloning to produce His-STAT3β_{TC} was performed by Jonathan Palmer (Kings College London).

The pET-28c(+) plasmid DNA vector was purchased from Novagen (Sutton, UK).

2.1.3 *Escherichia coli* Cells (*E. coli*)

XL1-Blue^{TM} *E. coli* competent cells **-**Genotype: *recAI, endAI, gyrA96, thi-I, hsdR17, supE44*, *relAI, lac* [F′ *proAB*$lacI^{q}$*ZΔM15* Tn*10* (Tet^{r})]. XLI-Blue *E.coli* cells were used for molecular cloning and optimal plasmid DNA amplification. This strain was purchased from Stratagene (Agilent Technologies, UK).

DH5α™ *E.coli* competent cells – Genotype: F^{-} ϕ80*lac*ZΔ15 Δ (*lac*ZYA-*arg*F) U169 *deo*R*rec*A1 *end*A1 *hsd*R17 (r_K^-, m_K^+) *pho*A*sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ^-. DH5α was employed for routine sub cloning application and high-quality DNA plasmid preparations. This strain was purchased from Invitrogen Ltd (Paisley, UK).

BL21 (DE3) TKB1™ *E.coli* competent cells – Genotype: *E.coli* B F^- DCM, ompT, hsdS (r_B^- m_B), gal λ (DE3) [pTK Tet^{r}]. TKB1 strain carries the gene for T7 polymerase and also contains a plasmid encoded, inducible tyrosine kinase gene (pTK). It was used for the controlled expression and phosphorylation of recombinant proteins placed downstream of the T7 polymerase binding site. This strain was purchased from Stratagene (Agilent Technologies, UK).

BL21 (DE3) Rosetta™ competent cells – Genotype: F- ompT, $hsdS_B$(r_B^- m_B) gal, dcm, $pRARE^{2}$ (Cam^{R}). This Rosetta™ strain enhances the expression of genes that encode six rare *E. coli* codons. It was used to overcome translational limitations by the codon usage of *E. coli*. It was used to express recombinant proteins with no post-translational modifications. This strain was purchased from Novagen (Merck bioscience, UK).

2.1.4 Buffers

All buffers were either autoclaved before use or passed through a 0.22 µm sterilizing filter in small volume.

2.1.4.1 Kinasing Medium

5X Modified M9 (1 L): 128 g $Na_2HPO_4.7H_2O$, 30 g KH_2PO_4, 5 g NaCl and 10 g NH_4Cl.

Compound	Preparation	Volume for 2L
5X Modified M9	Autoclaved	400 ml
1M $MgSO_4$	Autoclaved	2 ml
20% (w/v) D+ Glucose	Sterile filtered	20 ml
20% Casamino acids	Sterile filtered	10 ml
0.5% Thiamine HCl	Sterile filtered	2 ml
2.5 mg/ml Indole-3-acetic acid (IAA)	Sterile filtered	8 ml
50 mg/ml Ampicillin	Sterile filtered	2 ml
12.5 mg/ml Tetracycline	Sterile filtered	2 ml
DDW	Autoclaved	1554 ml

Table 2.4 Composition of TKB-1 kinase induction media

2.1.4.2 SDS PAGE Gels

Resolving gels composition

	7%	10%	12%	15%	18%
DDW	5.25 ml	3.75 ml	2.75 ml	1.148ml	0
1M TrisHCl (pH 8.8 at 4°C)	6 ml	6.0 ml	6.0 ml	6.0 ml	6.0 ml
Protogel	3.5 ml	5.0 ml	6.0 ml	7.5 ml	9.0 ml
10%, w/v, SDS	120 µl	120 µl	120 µl	120 µl	120 µl
10%, w/v, Ammonium persulphate	120 µl	120 µl	120 µl	120 µl	120 µl
TEMED	12 µl	12 µl	12 µl	12 µl	12 µl

Table 2.5: Resolving gel acrylamide percentage recipes

Stacking gel composition

Reagent	Volume
DDW	3.551 ml
1M TrisHCl (pH 6.8 at 4°C)	676 µl
Protogel	867 µl
10%, w/v, SDS	52 µl
10%, w/v, Ammonium persulphate	52 µl
TEMED	5.2 µl

Table 2.6: 5% stacking gel recipe

2.1.5 Antibiotics

Antibiotic stock concentrations and solvents (dilute 1/1000 for working concentration):

Antibiotic	Stock concentration (1000x) and solvent
Ampicillin	100 mg/ml in DDW
Tetracycline	12.5 mg/ml in 80% Ethanol
Kanamycin	30 mg/ml in DDW
Chloramphenicol	20 mg/ml in 80% Ethanol
Streptomycin	50 mg/ml in DDW

Table 2.7 Stock antibiotic preparations

2.1.6 Antibodies

STAT3 K-15 rabbit polyclonal IgG antibody raised against peptide region (626-640) of STAT3 of mouse origin was purchased from Santa Cruz Biotechnology (Wiltshire, UK). STAT3 (Phospho pY705) rabbit monoclonal IgG antibody raised against a synthetic phosphor-peptide corresponding to the residues surrounding the tyrosine 705 region of human STAT3 was purchased from AbCam (Cambridge, UK).

The phosphotyrosine mouse IgG monoclonal antibody (pY20) raised against phosphorylated tyrosine residues was purchased from BD Transduction laboratories (Oxford, U.K).

Anti-polyhistidine mouse IgG monoclonal antibody raised against synthetic polyhistidine tags was purchased from Sigma-Aldrich® (Dorset, UK).

ECL™ Peroxidase labelled Anti-Rabbit IgG horseradish peroxidase–linked whole antibody and ECL™ Peroxidase labelled Anti-Mouse IgG horseradish peroxidase–linked whole antibody were purchased from GE Healthcare (Buckinghamshire, UK).

2.1.7 Tissue Culture Reagents

MDA-MB-231 cells were grown in Dulbecco's modified Eagles medium (DMEM), GlutaMAX media supplemented with 5% foetal bovine serum (FBS), 5% minimum essential medium (MEM), non-essential amino acids (NEAAs), and 50 U/ml penicillin/streptomycin, all from Gibco by Life Technologies (Thermo Fisher).

A4 cells were grown in M^{c}Coy's 5A media (+$NaHCO_3$, -Glutamate)(Sigma) supplemented with 5% foetal bovine serum (FBS), 1% minimum essential medium (MEM), non-essential amino acids (NEAAs), 2mM L-glutamate (200mM), and 50 U/ml penicillin/streptomycin (5000 U/ml) all from Gibco by Life Technologies (Thermo Fisher).

NCI-H1975 cells were grown in RPMI 1640 media supplemented with 5% foetal bovine serum (FBS), 2mM L-glutamate (200 mM) and 50 U/ml penicillin/streptomycin (5000 U/ml) all from Gibco by Life Technologies (Thermo Fisher).

TrypLE express (1X) trypsin substitute was from Gibco by Life Technologies (Thermo Fisher), and Trypan blue solution (0.4%) and Thiazolyl Blue tetrazolium bromide came from Sigma-Aldrich (Poole, UK). Bambanker cell storage reagent was from Lymphotec Inc. (Tokyo, Japan).

Nunclon delta sterile tissue culture flasks (T25, T75 and T125ml) were from Nunc (Fisher Scientific Leicestershire UK), as were the Nunclon Delta surface 96-well sterile flat-bottomed assay plates. Fifty ml sterile reagent reservoirs came from VWR International Ltd (Dorset, UK) and 6-well sterile assay plates were from SPL life sciences (Pocheon, South Korea).

2.1.8 Assay Reagents

Eu^{3+} Cryptate-conjugated mouse monoclonal antibody (anti-6 Histidine) and d2-conjugated streptavidin were both purchased from Cisbio Bioassays

2.1.9 Equipment

A Dyad DNA engine was programmed for PCR, restriction digestion and general incubation. A New Brunswick Innova™ 4300 Innova™ rotatory incubator was used for bacterial culturing at 37°C; a Biochrom libra S22 spectrophotometer was used to estimate protein and DNA concentration; a Leec incubator, programmed at 37°C for overnight incubation; an Electrolab P300 fermenter, used for growing up to 10 L cultures; a Beckman Coulter Avanti J-26 XP centrifuge was used to harvest bacterial cell cultures; a MSE Soniprep 150 sonicator was used to lyse *E.coli*; a UVP GelDoc It^2 Imager was used to view and analyse stained DNA agarose and SDS protein gels, as well as taking images of Western blots; a Stuart gyro rocker SSL3 was used for incubation with shaking at either 4°C or at room temperature; An Ohaus pioneer and a Fisher Scientific SG2001 were used as basic balances, for routine weighing operations; Hanna HI 208 pH meter; Beckman Coulter centrifuge (XL90 Ultrafuge) for separating bacterial lysate into soluble and insoluble fractions at RCF greater than 20,000 (rotor used JA 25.5 rotor); Bibby B212 hot plate and stirrer, for heating or stirring buffer filtrations; Eppendorf table top centrifuge 5702R was used for low speed applications and protein concentration with the aid of Millipore Amicon and microcon concentration spin columns; Bio-Rad's Sub-Cell™ GT system was used for agarose gel electrophoresis; Bio- Rad Mini-Protean™ Tetra system used for SDS-PAGE and analysis; Bio-Rad's Mini-Protean II™ system used for Western Blots; AKTA purifier chromatographic system used for gel filtration chromatography; PerkinElmer Multilabel plate reader (Wallac

EnVision 2101) and Tecan i-control units were used to read 96 and 384-well assay plates at different excitation and emission wave lengths. Bench top vortexing was performed using an IKA MS3 basic, and static heating operations utilised a Grant-bio PCH-2 dry block heater. Double distilled water and molecular biology grade water were provided by ELGA purelab option and Ultra polisher units.

The tissue culture room consisted of a class II Envair Bio 2+ hood with aspiration provided by a Buchi V-700 vacuum pump; cells were incubated at 37°C in a Binder C150 at 5% CO_2 concentration. Cells were viewed using a Nikon eclipse TS100 microscope and reagents were brought to temperature with a Grant JB aqua 12 plus water bath and a Stuart Orbital SI50.

2.1.10 Peptides

The GP130 derived H-pYLPQTV-NH_2 and H-GpYLPQTV peptides and SH2 derived peptides H-pYLKTKF-NH_2, H-pYLKTKFI-NH2, and H-GpYLKTKFI used in the HTRF assay, along with scrambled variants, were either purchased from Cambridge Research Biochemicals Ltd. (Bellingham, UK) or Peptide Synthetic Protein Research Ltd. (Fareham, UK).

2.1.11 Consumables

Falcon polypropylene conical bottom tubes were purchased from BD Bioscience; 5, 10 and 25 ml serological pipettes, as well as Pasteur pipettes, were purchased from Greiner Bio-One; 20, 200, and 1000 µl micro pipette tips and gel-loading pipette tips were TipOne from Star Labs.

Hybond C and Hybond N+ nitrocellulose membranes were purchased from GE Healthcare (Buckinghamshire, UK).

Millex® GP syringe driven filter units (0.22 and 0.45 µM) were purchased from Millipore (Cork, Ireland). Syringes of various volumes were from Terumo (Surrey, UK). Parafilm was purchased from Pechiney plastic packaging. (Slough, Berkshire)

Superpose™ 12, agarose chromatography column, HiTrap™ QFF anion exchange columns and HISTrap FF and HP columns were purchased from GE Healthcare (Buckinghamshire, UK).

The TR-FRET assay was carried out in white, 384 well, small volume assay plates from Greiner bio-one.

2.1.12 Primers and Oligonucleotides

5'-BIOTINhexyl-TGCATTTCCCGTAAATCT-3'

5'-AAGATTTACGGGAAATGC-3'

The biotin labelled and unlabelled high affinity (*ds*M67) oligos used in the TR-FRET assays were synthesized by atdbio (Southampton, UK).

2.2 Biological Methods

All solutions were sterilized either using 0.22 µM syringe filters or autoclaved at 121°C for 25 mins prior to use.

2.2.1 *Escherichia coli* (*E. coli*) Stock Preparation and Transformation

Methods for *E. coli* cell competency preparation, transformation, storage and DNA extraction.

2.2.1.1 Preparation of Chemically Competent *E. coli* Cells.

A glycerol stock of *E.coli* was thawed on ice and, using a sterile loop, a streak of cells was transferred onto a Luria Bertani Broth (LB) 2.5% (w/v), agar 1.5% (w/v) plate supplemented with the appropriate antibiotics. The plate was inverted and incubated overnight (~16 hrs) at 37°C. A seed culture was then generated by adding a single colony from the plate to 10 ml LB broth (For 1 L): 10 g tryptone, 10 g NaCl, 5 g yeast extract (plus antibiotics) and incubated at 37°C for 4 to 6 hrs in a shaking incubator. One ml of this seed culture was then used to inoculate 500 ml sterile LB with antibiotics. The culture was incubated at 37°C while shaking at 150 rpm until an optical density (OD) $_{\lambda=600}$ = 0.6 was achieved. At an OD of 0.6 the cells were in late-log phase of growth and still metabolically active. The culture was transferred to a pre-chilled 1 L centrifuge pot and cooled on ice for 10 mins. The culture was then centrifuged at 3600 rpm for 20 mins at 4°C. The supernatant was discarded and the pellet re-suspended in 50 ml ice-cold sterile 100 mM $CaCl_2$. Maintaining the solution on ice, the re-suspended cell suspension was diluted further to 250 ml with ice-cold 100 mM $CaCl_2$ and cooled for 20 mins. The culture was then centrifuged again at 3600 rpm for 25 mins at 4°C. The supernatant was once again discarded and the pellet re-suspended in 5 ml ice-cold $CaCl_2$ storage solution (20% (w/v) glycerol, 100 mM $CaCl_2$). The cell suspension was

immediately aliquoted into ~25 x 200 µl fractions, snap frozen in liquid nitrogen and stored at –80°C until use.

2.2.1.2 *E. coli* Competent Cell Transformation

Competent *E. coli* was used for vector amplification and antibiotic resistance selection. An aliquot of frozen competent *E.coli* cells was thawed on ice and ~0.50 µg (1 µl) of plasmid DNA added. The sample was then gently mixed and incubated on ice for 45 mins. The cell-plasmid mixture was then transferred to a heating block to perform the heat shock stage (42°C for 90 secs). The cells were then immediately returned to the ice for a further 2 mins to recover. A further 400 µl of sterile LB broth was then added to the sample and the cells were grown with agitation (150 rpm for 45 mins at 37°C). After incubation, 100 µl of the culture media was spread on the appropriate antibiotic selective plate under sterile conditions. The plates were inverted and incubated at 37°C for 16 hrs. Bacterial colonies were either selected for further studies or the plates were wrapped in parafilm and stored at 4°C for future use.

2.2.1.3 Glycerol Stock of Transformed *E. coli* Cells

Freshly prepared sterile LB broth (5 ml) supplemented with the appropriate antibiotics was inoculated with a single colony from a plate of transformed *E. coli* cells. The culture was grown at 37°C with agitation (150 rpm) until an $OD_{\lambda 600}$ of 0.6 was reached. Glycerol 50% (w/v) stocks of the culture media were prepared by diluting 500 µl culture with 500 µl sterile glycerol. The mixture was thoroughly mixed by inversion and stored at -20°C until use.

2.2.1.4 Mini Preparation of Plasmid DNA

Five ml of freshly prepared sterile LB broth was transferred into a 50 ml Falcon tube and then inoculated with a single colony of transformed *E. coli* picked from a selected plate using a sterile loop. The 5 ml culture was supplemented with the appropriate antibiotics, as indicated in table 2.7. The culture was grown at 37°C with shaking at 150 rpm overnight. 1.5 ml of the overnight culture was transferred into a sterile 1.5 ml Eppendorf tube and then centrifuged at 7000 RCF for 2 mins at 4°C. The resulting supernatant was discarded and the cell pellet was re-suspended in 200 µl of re-suspension buffer {50 mM TrisHCl (pH 8.0 at 4°C), 10 mM EDTA, 100 µg/ml RNase A) followed by 200 µl of lysis buffer {200 mM NaOH, 1% (w/v) SDS}. The mixture was then inverted gently 6 times. The lysate was then immediately neutralised with 200 µl of neutralization buffer (3M potassium acetate, pH 4.8), and mixed gently again by inverting the tube 6 times until a white precipitate developed. The tubes were then incubated at -20°C for 10 mins and the precipitate pelleted by centrifugation at 13,000 RCF for 10 mins at 4°C. The resulting supernatant was transferred into a fresh 1.5 ml Eppendorf tube with 400 µl of ice-cold isopropanol and mixed by inversion, followed by a further centrifugation at 13000 RCF for 10 mins at 4°C. The supernatant was discarded, 200 µl of ice-cold 70% ethanol (w/v) was added and mixed to wash the near colourless DNA pellet. The tubes were once more centrifuged at 13,000 RCF for 5 mins at 4°C and the supernatant discarded. The plasmid DNA pellet was air-dried for 10 mins prior to suspension in 50 µl of sterilized ddH20. The sample was stored at -20°C for future use.

2.2.1.5 Maxi Preparation of Plasmid DNA

The Qiagen Plasmid Maxiprep kit was employed for large scale plasmid DNA extraction and purification. A single colony of a transformed *E. coli* cell picked from an antibiotic selective plate was used to inoculate 5 ml of freshly prepared sterile LB broth supplemented with the

appropriate selective antibiotics (see Table 2.7). The culture was grown overnight at 37°C shaking at 150 rpm. The 5 ml culture was used to inoculate 500 ml of freshly prepared sterile LB broth supplemented with antibiotics. The culture was further incubated overnight at 37°C with shaking at 150 rpm, followed by centrifugation at 3600 rpm for 20 mins at 4°C.

Using the Qiagen Maxi prep protocol, the cell pellet was re-suspended in 10 ml of re-suspension buffer (P1) by gentle pipetting. Ten ml of lysis buffer (P2) was added and mixed gently by inversion 6 times and then incubated for 4 mins at room temperature. Ten ml of ice cold neutralization buffer (P3) was added to stop lysis. The mixture was thoroughly mixed by inversion until a white precipitate was formed and the resulting mixture centrifuged at 20,000 relative centrifugal force (RCF) for 30 mins at 4°C.

The supernatant was then passed through the equilibrated (equilibration buffer; buffer QBT) Qiagen-tip column by gravity flow. The column was washed twice with 30 ml buffer QC and then eluted with 15 ml buffer QF. The eluted fraction containing the plasmid DNA was subsequently precipitated with 10.5 ml of isopropanol and immediately centrifuged at 20,000 RCF for 30 mins at 4°C. The resulting precipitated DNA pellet was washed twice with 70% ethanol to remove salts, as well as to substitute isopropanol with more volatile ethanol as this increases the DNA's solubility. The ethanol-DNA suspension was centrifuged at 20,000 RCF for 30 mins at 4°C and the plasmid DNA pellet air-dried for 10 mins before re-dissolving in 500 µl of sterilized double distilled H_2O.

The concentration of the purified plasmid DNA was estimated by reading the ODλ = 260nm. Where an ODλ = 260nm of 1.0 is equivalent to a DNA concentration of 50 µg/ml. The plasmid DNA preparation was evaluated by agarose gel electrophoresis after restriction endonuclease digestion to ensure that the correct plasmid had been amplified and purified. The sample was stored at -20°C.

2.2.2 Unphosphorylated HIS-STAT3β$_{TC}$ Production and Purification

Methodologies for expression, extraction and purification of recombinant proteins.

2.2.2.1 Expression of Unphosphorylated HIS-STAT3β$_{TC}$

Five hundred ml of LB broth was prepared (12.5 g Luria Base broth with 500 ml DDW), added to a 2 L shake flask and sterilized at 121°C for 20 mins. Ten ml of sterile LB in a sterile 20 ml container containing the required antibiotics was inoculated with 10 µl of master cells. This seed culture was then incubated overnight in a shaking incubator at 37°C. The following day the 2 L shake flask was pre-warmed to 37°C. Tetracycline and kanamycin were added to the flask to give the working final concentrations (see appendices). The flask was inoculated with 10 ml of the overnight seed culture and incubated in the shaker at 37°C and >150 rpm. Once the OD_{600} nm reached 0.6AU (late logarithmic growth phase) the culture was induced with 500 µl of 1M IPTG to give a final concentration of 1 mM IPTG, the incubator was then reduced to 21°C and left shaking overnight. IPTG acts as a molecular mimic of allolactose, a metabolite that initiates transcription of the *lac* operon by releasing a tetrameric repressor from the lac operator. It is used to induce protein expression where the gene of interest is under the control of the lac operator, as in the pET expression vector system.

The culture was harvested into pre-weighed 1 L centrifuge pots and centrifuged at 3600 x *g* for 20 mins at 4°C. The pellet was re-suspended in 30 ml/g of supernatant and then re-centrifuged at 3600 x *g* for 20 mins at 4°C in 50 ml centrifuge tubes resulting in approximately 1 g cell pellets. The supernatant was discarded and the pellets frozen until required.

2.2.2.2 Extraction of Unphosphorylated HIS-STAT3β$_{TC}$

One hundred ml of STAT3 extraction buffer (20 mM HEPES-KCL pH 7.6, 10% glycerol w/v, 1 mM EDTA, 10 mM $MnCl_2$, 20 mM DTT, 0.5 mM PMSF and 1 x SigmaFast protease inhibitor tablet) was freshly prepared. HIS-STAT3 cell pellets were removed from -20°C storage and left to defrost on ice. The pellets were re-suspended in 30 ml/g of extraction buffer using either a potter homogenizer or serological pipette. The re-suspended cell pellet was placed on ice (to minimize temperature variations) and sonicated on a 15 secs on and 15 secs off cycle for 5 mins with an amplitude of 15 microns. The resulting post sonication lysate was then centrifuged at 27,000 x *g* for 1 hr at 4°C. The supernatant containing HIS-STAT3 was decanted into a 100 ml beaker and stirred at 4°C whilst 7.5 g (1.9 M final concentration) of ammonium sulphate was gradually added to facilitate the precipitation of the protein. Ammonium sulphate is used to 'Salt-Out' protein by increasing the ionic concentration of the solution to the point where the HIS-STAT3 precipitates. Once all the ammonium sulphate was added the solution was left to stir for 20 mins at 4°C. Following the incubation, the material was centrifuged at 30,000 x *g* for 1 hr at 4°C. The resulting protein pellet containing HIS-STAT3 was retained for further anion exchange chromatographic purification.

2.2.2.3 Purification of HIS-STAT3β$_{TC}$

Further purification of HIS-STAT3β was conducted by anion exchange chromatography, in which the HIS-STAT3β and contaminant proteins carry a net negative charge at pH 7.6 (as this pH is more basic than the isoelectric point of HIS-STAT3β pI = 7.58) and, therefore, binds to the positively charged column media. This bound protein mixture is then eluted by the addition of a sodium chloride gradient which then fractionates the proteins based on their

affinity for the column. This is because different proteins have differing isoelectric points and thus have slightly different affinities for the column, which, when a salt gradient is applied, results in the fractionation of the sample, and the purification of STAT3.

The QFF column was first washed with 2 CV (10 ml) of salt free buffer at a flow rate of ≤1ml/min to remove the 20% w/v ethanol storage solution from the column. Then the column was cleaned with 1M NaCl in ion exchange buffer for 4 CV at ≤1ml/min, to remove any bound contaminants. Lastly, the column was re-equilibrated in salt free ion exchange buffer (100 mM TrisHCl (pH 8.5 at 4°C), 1 mM EDTA, 2 mM DTT), and was ready for loading.

The STAT3 extracted salt pellet was thoroughly re-suspended in 5 ml of salt free ion exchange buffer. This was then passed through a 0.22 µm filter, and then further diluted to 20 ml using DDW. This material was then ready for loading onto the QFF column, and is now referred to as the starting material.

2.2.2.4 Ion Exchange Chromatography

Five ml of the STAT3 starting material was applied to the prewashed QFF column at a low flow rate of ≤0.5 ml/min to allow the protein to bind. The flow through was collected and the last drop to leave the column was added to 1 ml of Bio-Rad protein reagent in an Eppendorf tube and compared with a reference standard (1 ml Bio-Rad +1 drop of salt free ion exchange buffer). This was checked by eye to see if a significant amount of protein was not binding to the column. If little or no protein was detected in the flow through then another 5 ml was loaded onto the column and checked as before using the Bio-Rad protein reagent. Once protein was no longer binding to the column, as seen in the flow through, the column could be washed. The column was washed with 2 CV of salt free ion exchange buffer, which was collected. Elution was carried out by flowing 0.1M NaCl in ion exchange elution buffer {100 mM TrisHCl (pH. 8.5)} at 4°C, 1 mM EDTA, 2 mM DTT, 0.1 M NaCl – 1M NaCl) through

the column at ≤0.5ml/min, and collecting 1 ml fractions; this is predominantly when STAT3 elutes. Then, 2 CV of 0.2 M NaCl in ion exchange buffer was flowed through the column and was collected in a 10 ml tube. This was then repeated using 0.3 M NaCl and 0.4 M NaCl.

This whole procedure was repeated until all the STAT3 starting material passed through the column and was eluted. A Bio-Rad protein assay was then performed on all the fractions collected. SDS-PAGE and Western blot analysis was performed on the 0.1 M elution samples and some other representative samples, as required.

2.2.2.5 Affinity Chromatography

A 5 ml HIS-Trap FF was washed with four column volumes of DDW in order to remove the ethanol storage solution and then equilibrated with four column volumes of equilibration buffer (20 mM sodium phosphate, 0.5 M NaCl, 20 mM imidazole, pH 7.4). The protein solution was diluted to 20 ml and passed through a 0.45 µm syringe filter and then slowly (<1ml/min) passed through the HIS-Trap column whilst collecting the flow through. The flow through was again passed through the column to ensure maximum binding efficiency; again the flow through was retained to run on SDS gel.

The column was then washed with 20 column volumes (100 ml) of equilibration buffer at 2 ml/min in order to remove any impurities. The bound protein was eluted by the slow addition (0.5 ml/min) of 10 ml of elution buffer (20 mM sodium phosphate, 0.5 M NaCl, 500 mM imidazole, pH 7.4) and the eluate was collected in 1 ml fractions and analysed by SDS-PAGE and dialysed into the desired buffer.

2.2.3 Production of Phosphorylated HIS-STAT3β$_{TC}$.

Methodologies for expression, extraction and purification of recombinant proteins.

2.2.3.1 Transformation into BL21 (DE3) TK Cells

Two hundred µl of TKB1 competent cells were thawed on ice and mixed by pipette. One hundred µl of cells were aliquoted into two pre-chilled 14 ml BD Falcon polypropylene round-bottom tubes. A diluted final concentration of 25 mM β-mercaptoethanol was added to the 100 µl of competent cells, in order to increase transformation efficiency. The contents of the tube were swirled gently and incubated on ice for 10 mins, swirling the tube gently every 2 mins. Two µl of HIS-STAT3β-TC DNA (~50 ng/µl) was added to the polypropylene tube and swirled gently. As an optional transformation efficiency control, 1 µl of pUC18 control plasmid was added to the other 100 µl aliquot of TKB1 cells and swirled gently. The two polypropylene tubes were incubated on ice for 30 mins, while sterile LB media was pre-warmed in a 42°C water bath. The tubes were then placed in the 42°C water bath for 45 secs and subsequently plunged back on ice for 2 mins. The pre-warmed LB broth (0.9 ml) was added to each of the tubes, which were then placed in a shaking incubator for 1 hr (37°C, 225-250 rpm). One hundred µl of HIS-STAT3β-TC transformed TKB1 cells were plated on LB-tetracycline-kanamycin plates (LB-tetracycline-ampicillin plates for the pUC18 transformed control cells) using a sterile spreader and grown at 37°C overnight.

2.2.3.2 Expression of Phosphorylated HIS-STAT3β$_{TC}$

Six L of LB broth was prepared (150 g Luria Base broth with 6000 ml DDW), added to a 10L fermenter, and sterilized at 121°C for 20 mins. Ten ml of sterile LB in 20 ml sterelin containing the required antibiotics was inoculated with 10 µL of glycerol stock master cells. This seed culture was then incubated overnight in a shaking incubator at 37°C. The following

day the fermenter was pre-warmed to 37°C. Tetracycline and kanamycin were added to give the working final concentrations (see appendices). The fermenter was inoculated with 10 ml of the overnight seed culture, set to incubate at 37°C and stirred at 150 rpm. If excess foaming occurred, then a single drop of antifoam was added (Sigma Cat No. A-6426). Once the OD_{600} nm reached 0.6 AU the culture was induced with 6 ml of 1M IPTG to give a final concentration of 1 mM IPTG, the temperature was reduced to 21°C, and left overnight.

The cells were harvested the following morning by peristaltic pump and centrifuged at 3600 x *g* for 25 mins at 4°C to obtain a cell pellet that was subsequently re-suspended in a small amount of kinasing medium. The indoleacrylic acid present in the media induces the expression of the Elk receptor protein-tyrosine kinase domain under the control of the trp promoter (170). The remaining kinasing medium was added to the rinsed fermenter, along with the re-suspended pellet, which was heated to 37°C and left mixing for 2.5 hrs.

The culture was harvested into pre-weighed 1 L centrifuge pots and centrifuged at 3300 x *g* for 25 mins at 4°C. The pellet was re-suspended in 30 ml/g of supernatant and then re-centrifuged at 3600 x *g* for 20 mins at 4°C in 15 ml falcon tubes, resulting in approximately 1 g cell pellets. The supernatant was discarded and the pellets frozen at -20°C until required.

2.2.3.3 Extraction of Phosphorylated HIS-STAT3β$_{TC}$

pHIS-STAT3βTC was extracted using the same method for unphosphorylated HIS-STAT3βTC, as described in section 2.2.2.2; however, the post sonication lysate was progressed to purification, as described in 2.2.4, without ammonium sulphate precipitation.

2.2.3.4 Purification of Phosphorylated HIS-STAT3β_{TC}

HIS-P-STAT3β_{TC} was purified primarily by ion exchange chromatography; the STAT protein carries a net negative charge and therefore binds to the positively charged column. The bound protein is then eluted by raising the sodium chloride concentration through a gradient. Our protein of interest is eluted when the isoelectric point is reached and therefore we achieve fractionation and purification of the proteins.

The QFF anion exchange column was prepared by washing it with 2 column volumes (10 ml) of salt free buffer at a slow and steady rate (<1 ml/min) in order to fully remove the 20% ethanol column storage solution. The column was then flushed with four column volumes of 1M NaCl ion exchange buffer to ensure that the column was cleaned of all contaminants. Finally, the column was re-equilibrated with four column volumes of salt-free ion exchange buffer ready for sample loading.

Extracted pHIS STAT3β_{TC} was filtered through a 0.45 µm particulate filter and diluted to 20 ml with salt-free ion exchange buffer. The diluted protein solution was then applied to the equilibrated QFF column at a low flow rate of ≤ 0.5 ml/min to maximise protein binding. Flow through was collected and re-applied to the column; the flow through was collected for SDS-PAGE analysis. The column was then washed with two column volumes of salt free ion exchange buffer, which was collected for SDS-PAGE. Elution of the protein was carried out by adding 0.1M NaCl in ion exchange buffer through the column at ≤0.5 ml/min, and collecting 1 ml fractions (most P-HIS-STAT3β_{TC} is eluted at this concentration). Then 2 CV of 0.2M NaCl in ion exchange buffer was flowed through the column and collected in a 10 ml tube. This was then repeated using 0.3 M NaCl and 0.4 M NaCl. A Bio-Rad protein assay was then performed on all collected fractions and SDS-PAGE/Western blot analysis, as required.

2.2.4 Homogenous Time-Resolved FRET STAT3β_{TC} Assay

Reagent preparation and addition priority methodologies.

2.2.4.1 Annealing of Oligonucleotide DNA Primers

An annealing mixture was prepared by mixing equimolar amounts of complementary DNA in a sterile 1.5 ml tube (*e.g.* 20 µl forward and reverse primers of 100 µM each). The reaction mixture was supplemented with 100 mM NaCl, and then diluted to 100 µl with sterile ddH2O. The lid of the 1.5 ml tube was wrapped in parafilm and then supported in a water-bath. The water-bath was heated until boiling (100°C) for 5 mins and subsequently allowed to cool slowly to room temperature. The tube was centrifuged for 30 secs and the oligonucleotide DNA stored at -20°C for future use.

2.2.4.2 Preparation of Fluorescent Conjugates

The Streptavidin-d2 and Anti-6HIS-Cryptate reagents (Cisbio Bioassays) were supplied as lyophilised 250 µg samples. These were reconstituted with 250 µl of sterile DDW to give a 1 mg/ml stock. This stock was aliquoted into 25 x 10 µl samples, which were frozen at -20°C until use. A 10 µl aliquot was thawed on ice when required and diluted with 990 µl of TR-FRET binding buffer {100 mM TrisHCl (pH. 7.4) at 4°C, 10 mM sodium phosphate, 0.1% (w/v) BSA, 0.1% (w/v) Tween 20, 0.5% (w/v) glycerol}. Five µl was used per test well (50 ng/well) and once prepared the conjugates were kept either on ice or at 4°C and disposed of after two weeks.

2.2.4.3 HTRF Assay Setup

The concentration of purified HIS-STAT3β_{TC} was determined by Bradford assay (as in section 2.2.10.5). This concentration in mg/ml was converted to Moles by dividing the concentration by the proteins molecular weight (HIS-USTAT3β_{TC} = 70392Da, HIS-PSTAT3β_{TC} = 70487Da). The protein was then diluted with TR-FRET buffer to the working concentration.

The diluted protein was incubated with the DNA as a stock at 4°C for 4 hrs (in order for an equilibration between protein and DNA to be established) prior to aliquoting 10 µl per well into the required number of plate wells. One µl of inhibitor prepared in DMSO was then added to the test wells and gently pipette mixed. The inhibitors/controls were incubated on a plate shaker (300 rpm) at 4°C for 1 hr and then 10 µl of the 1:1 Streptavidin-d2 and Anti-6HIS-Cryptate fluorophores mixture was added, pipette mixed and again incubated on the plate shaker at 4°C for 1 hr. One µl of 10M KF was added and pipette mixed in all wells prior to reading the plate.

The Envision plate reader was set up to excite at a wavelength of 320nm and emission readings were recorded at 615 nm and 665 nm after a 60 µs delay with 100 flashes.

2.2.5 Cell Culture

Methodologies for cell passage and counting, culture maintenance and assays.

2.2.5.1 General Procedure for Cell Culture Passage

Cultures were expanded as required. Cell cultures were split (passaged) every 2–4 days, when they reached 70–80% confluence.

The following passage protocol is described for the standard T75 flask. Media was aspirated off and the cells washed once with 10 ml pre-warmed sterile PBS (containing no Ca2+ or Mg2+). Two ml of trypsin-EDTA solution was added and incubated for 5 mins in the incubator until the cells had detached. To cease trypsinisation, 10 ml of growth medium was added and the resulting cell mixture transferred to a sterile falcon tube and centrifuged for 5 mins at 1200 rpm in a pre-cooled (4°C) centrifuge. The cell pellet was then re-suspended in 5 ml of media and the cells counted (as in Section 2.2.6.2). A newly labelled T75 had 14 ml of fresh media added and 1 x 10^6 cells diluted to 1 ml. The flask was gently moved to evenly distribute the cells on the surface.

2.2.5.2 Cell Counting by Haemocytometer

Ten µl of cell suspension was added to 90 µl of Trypan blue solution and well pipette mixed. Ten µl of this mixture was added to the clean haemocytometer utilizing capillary action to draw the fluid beneath the cover slip. The four corner cells (each containing 16 squares) were counted (using a hand tally counter) under a 10X objective microscope. Only healthy unstained cells were counted and only cells within the square and any positioned on the right hand or bottom boundary line were included. The haemocytometer is designed so that the number of cells in one set of 16 corner squares is equivalent to the number of cells x 10^4 /ml

Therefore, to obtain the count:

The total count from 4 sets of 16 corner = (cells / ml x10^4) x 4 squares from one haemocytometer grid

1. Divide the count by 4

2. Then multiply by 10 to adjust for the 1:10 dilution in Trypan blue

2.2.5.3 General Procedure for Cell Frozen Stock Preparation

The cells were first trypsinised from the desired number of flasks and the cell suspensions pooled together whilst the cells were in the logarithmic growth phase. The cells were counted, and total viable cell number calculated. The cells were centrifuged at 125 x *g* for 10 mins and the supernatant aspirated.

The pellet was re-suspended at a density of 2 x 10^6 cells/ml in Bambanker freezing medium. One ml aliquots were dispensed into sterile cryovials, then frozen at -80°C without preliminary freezing. Two or more weeks later, the viability of the frozen stocks was confirmed by starting a fresh culture.

2.2.5.4 General Procedure for Cell Culture from Frozen Stock

Frozen stocks of cells were thawed rapidly by briefly immersing the vial in a 37°C water bath (e.g. 2–3 mins with constant agitation). Once thawed, the outside of the vial was immediately wiped with 70% EtOH and the contents of the vial transferred to a T25 flask.

An additional 4 ml of appropriate pre-warmed medium was then added to the flask. The flask was gently mixed to distribute the cells evenly over the growth surface. The culture was then placed in a 37°C, 5% CO_2, humidified incubator.

The next day, the cells were examined under a microscope to check for adhesion to the plate and normal morphology. Medium was aspirated and replaced with fresh.

2.2.5.5 General Procedure for the MTT Assay

10,000-20,000 cells in 100 µl media were plated into a Nunc Nunclon flat bottomed clear 96-well plate, and incubated for 24/48 hrs at 37°C in a 5% CO_2 incubator. One µl of compound dissolved in DMSO was added in triplicate along with a DMSO only control and the cells incubated for a further 24/48 hrs. After the required incubation the media was removed using a multichannel pipette and each well washed with 200 µL DMEM (Phenol red free), which was again removed with a multichannel pipette. A stock 5 mg/ml Thiazolyl Blue Tetrazolium Blue (MTT) (Sigma M5655) solution in DMEM (phenol red free) was prepared and 100 µl of a subsequent 0.5 mg/ml MTT stock was added to each well and left in the incubator for 4 hrs for the formazan crystals to develop. The reagent was then carefully removed, so as not to disturb the crystals, which were dissolved in 200 µl of DMSO (Sorenson's glycine buffer) and incubated to ensure that no air-bubbles remain.

Absorbance readings were taken at 570 nm (Envision plate reader) and assumed to be directly proportional to the number of living cells. Absorbance values for cells treated with either vehicle or media alone were used as controls. The MTT data (calculated as percentage of control values) were calculated from triplicate measurements to allow standard deviations to be calculated.

2×10^5 HeLa cells were plated onto 96-well plates with 100 µl DMEM media and left overnight. Media was removed and replaced the following morning and left for 8 hrs after which the media was replaced and inhibitors added with a serial 1:10 dilution to achieve titration along with media only and DMSO only controls. The cells were left at 37°C for 24 hrs.

2.2.5.6 General Procedure for Trypan Blue Assay

MDA-MB-231 cells were plated into 24-well plates and allowed to incubate overnight to achieve 80% confluency. Inhibitors were then added, and the cells incubated for a further 24 hrs. After addition of Trypan Blue, unstained (viable) and stained (non-viable) cells were counted and calculated as a percentage of total cells using a haemocytometer (as described in section 2.2.5.2).

2.2.6 *In Silico* Modelling

All modelling studies were carried out with the expertise of Dr Paul Jackson.

2.2.6.1 Receptor Preparation

The STAT3 homodimer structure was downloaded from the Protein Data Band (PDB ID: 1BG1) and missing residues and loop structures were generated using the Swiss Model online tool (http://swissmodel.expasy.org/interactive).

The STAT3 homodimer was then subjected to a number of steps of preparation in Chimera. Firstly, the receptor was assigned partial charges (AMBER ff98SB), after removal of hydrogens, using the DockPrep module of AMBER, followed by writing to mol2 and pdb files. A molecular surface of the receptor was generated using write dms. This process rolls a water molecule-sized ball over the surface of the receptor and it represents the van der Waal's surface of the molecule. Advantages of this step are that it helps to include deep crevices inaccessible to solvent molecules in later docking calculations and concurrently removes tiny pockets from the analysis which would be inaccessible to ligands.

2.2.6.2 Ligand Preparation

All ligands were constructed and energy minimised using ChemBioOffice and exported in pdb format. All files were then converted into the Sybyl format mol2 using Chimera (171). The DockPrep module in Chimera was then used to assign partial charges of each molecule (AM1-BCC charges in this instance). As part of the DockPrep procedure, atom types were subsequently assigned via the AMBER GAFF force-field using ANTECHAMBER.

2.2.6.3 Docking

All docking experiments were undertaken by Dr Paul Jackson using the DOCK6 software suite. The STAT3 monomeric crystal structure (PDB ID: 1BG1) was used in every experiment. In each case, homodimer A was used as the receptor and the PYLKTKFI peptide motif of homodimer B (created through unit cell generation in Chimera) was considered the ligand.

The receptor was prepared (outlined above), and a number of steps were undertaken in DOCK6 to isolate the binding pocket of interest. Firstly, spheres were generated around the surface of the molecule using sphgen, which produces sets of overlapping spheres to describe the surface of the receptor.

Sphere_selector was then used to filter results from sphgen, and only spheres within 10Å of the PYLKTKFI peptide were selected for further analysis. This resulted in the assessment of the full SH2 domain of the STAT3 molecule (residues 582 – 688) for potential binding of the KSN series of molecules.

A further identical study was conducted on the DNA binding domain (residues 321 – 465) using spheres within 10Å of residues 324–328 (VERQP) in order to ascertain the potential of the KSN series of molecules to bind to the DNA binding domain of STAT3.

Finally, the KSN series of molecules was automatically positioned into the spheres with maximum number of conformations set at a high level (500) to explore a large amount of conformational space, thus producing a library of docked ligand: protein structures.

2.2.6.4 Evaluation of Ligand Binding

Ligands were evaluated based on two DOCK scoring functions, MMGBSA and GRID scoring. During the docking process, a grid was created around the receptor. The grid was then used to allow rapid score evaluation in DOCK. Prior to scoring, orientations of the ligand which exhibited significant steric interactions with the receptor molecule were discarded using the bump filter.

The grid-based scoring term is based on non-bonded parameters within the force-field (i.e. van der Waal's forces, electrostatic forces). The Hawkins MMGBSA method (172) is an adaptation of the original MMGBSA method and uses the pairwise solvation method and the interaction between ligand and receptor are represented by Lennard Jones and Coulombic potentials, coupled with the change in solvation (ΔGBSA), which in turn is represented by the following equation:

$$\Delta GBSA = GBSA_{complex} - (GBSA_{receptor} + GBSA_{ligand})$$

In evaluating ligand-protein interactions, two factors were considered crucial; shape-fit of the molecule to the protein receptor, and prevention of the interaction of the STAT3 homodimer B (particularly residues PYLKTKFI) with an octapeptide binding site. The latter was considered particularly relevant as modulation of either Y705 (which is phosphorylated before dimerisation) or P704 (i.e. P and Y of PYLKTKFI) is known to disrupt dimerisation of the protein.(173)

The former was considered relevant as shape-fit is crucially important to the binding of other STAT3 binding molecules that target the SH2 domain of the protein (for example BP-1-102) (139).

2.2.7 Polymerase Chain Reaction

Methodologies for RNA isolation, reverse transcription PCR and endpoint PCR.

2.2.7.1 Cell Stimulation and RNA Isolation

The cancer cell line to be investigated was split and counted in the normal manner and a 6-well plate seeded with 100,000 cells in a 2 ml total volume of media. The cells were left overnight to establish themselves and the next day were stimulated with either lipopolysaccharide (LPS) or interleukin-6 (IL6). LPS (300 µg) was added to each well to give 150 µg/ml final concentration, or 8 µl per well of a 10 µg/ml IL6 stock was added to give a 40 ng/ml final concentration. These stimulants of the STAT pathway were left for 24 hrs before the introduction of either test compounds or control DMSO. The inhibitors were left overnight.

A cell pellet was prepared by carefully aspirating the media of the wells then washing the wells with 3 ml of PBS (8 g/l sodium chloride, 137 mM), 0.2 g/l potassium chloride (2.7 mM), 1.15 g/l di-sodium hydrogen phosphate (4.3 mM), 0.2 g/l potassium dihydrogen phosphate (1.47 mM, adjusted to pH 7.4). Trypsin-EDTA (0.2 ml) was added and, after ensuring an even covering, incubated at 37°C for 2 mins. After trypsinising the cells 2 ml of media was added and the cells transferred to a 15 ml Falcon tube, where a further 8 ml of media was added before centrifuging at 1500 rpm at 4°C for 5 mins. The media was carefully aspirated and the pellet re-suspended with 10 ml of PBS. The mixture was again centrifuged at 1500 rpm at 4°C for 5 mins and the PBS carefully aspirated. The cell pellet was snap frozen using liquid nitrogen and stored at -80°C until further use.

2.2.7.2 RNA Isolation

The Qiagen RNeasy plus mini kit was used to purify total RNA from the pelleted *E-coli* cells. Before use, the ethanol wash buffer (RPE) was prepared by diluting with 4 volumes of ethanol (96-100%) to working concentration and 350 µl added to the pre-thawed, on ice, cell pellet. The pellet was well homogenized using a pipette and vortex mixed. The lysate was then centrifuged in a minifuge for 3 mins and the supernatant carefully pipetted off and added to the genomic DNA eliminating spin column. The column was centrifuged at 15,000 rpm for 30 secs, and the flow-through retained. To the flow-through, 350 µl of 70% ethanol was added, mixed well with a pipette, and immediately transferred to the RNeasy spin column and centrifuged at 15,000 rpm for 15 secs; the flow-through was discarded. RNA wash buffer (RW1) (700 µl) was added to the spin column and centrifuged at 15,000 rpm for 15 secs; the wash buffer flow-through was discarded.

Then, 500 µl of ethanol wash buffer (RPE) was added to the spin column and centrifuged at 15,000 rpm for 15 secs before adding a second 500 µl to the column and centrifuging for 2 mins at 15,000 rpm and discarding the flow through. The column was centrifuged at full speed for 1 min more to ensure that the membrane was dry. A new collection tube was fitted and 50 µl of RNase free water carefully added to the centre of the spin column membrane. The column was centrifuged at 15,000 rpm for 1 min in order to elute the RNA.

2.2.7.3 Reverse Transcription Polymerase Chain Reaction (RT-PCR)

This step produces a cDNA version of the RNA present so it is important that these steps are carried out under sterile conditions and without RNase contamination.

PCR tubes were prepared containing 10 µl of reverse transcription buffer, 1 µl of reverse transcriptase enzyme mix and 9 µl of RNA sample. A negative control was prepared with PCR grade H_2O in place of the enzyme mix, giving all samples 20 µl total volume. The samples were well mixed by pipetting up and down and vortexing for 5 secs, followed by a short spin to ensure that all the sample was at the bottom of the tube.

A simple PCR program of 37°C for 1 hr., followed by 95°C for 5 mins, then held at 4°C was run on all samples. Samples could be stored at -20°C before endpoint PCR was carried out.

2.2.7.4 Endpoint PCR

A reaction master mix was prepared as in table 2.8

Components	Volume required (µl) for one reaction
PCR reaction buffer (10X)	2.5
dNTP mix (10 mM)	0.5
$MgCl_2$ (50 mM)	0.75
Forward primer (1:10)	0.625
Reverse primer (1:10)	0.625
Taq polymerase	0.1
PCR grade water	18.9

Table 2.8 Endpoint PCR master mix

All components were thawed on ice and vortex mixed before use (although the Taq polymerase was not vortexed). Autoclaved PCR tubes were used to prepare the master mix, and samples prepared straight away with 1 µl of cDNA sample, followed by 24 µl of master mix, and pipette mixed; a control sample containing water instead of Taq polymerase was also prepared.

The samples were placed in the PCR thermo cycler and a program run as follows

Stage 1	94°C for 5 mins
Stage 2	94°C for 30 secs
	55°C for 30 secs
	72°C for 90 secs
Stage 3	Repeat stage 2 for 30 cycles
Stage 4	72°C for 5 mins

All samples were subsequently run on an agarose gel.

2.2.8 Cell Based Western Blot

1 x 10^6 cells were plated into 2% FCS media in 6-well plates and incubated overnight to initiate attachment and differentiation. The medium was then changed to a serum-free type, and either IL-6 (20 ng/ml) or lipopolysaccharide (500 µg/ml) was added at the same time as the inhibitor. Whole cell extracts were then prepared using RIPA buffer (Thermo Scientific) with protease and phosphatase inhibitors (1.6 mg/ml benzamidine HCl, 1.0 mg/ml phenanthroline, 1.0 mg/ml aprotonin, 1.0 mg/ml leupeptin, 1.0 mg/ml pepstatin A dissolved in 99% ethanol). Extracts were dissolved in SDS buffer and run on a 10% PAGE gel for 1 hr at 100 V (4°C). Blots were probed with antibodies for P-STAT3, STAT1, Cyclin D1, Survivin and GAPDH (Cell Signalling Inc.).

Once treated the cell culture dish was placed on ice, carefully aspirated and each well washed with 2 ml ice-cold PBS. (8 g/l sodium chloride (137 mM), 0.2 g/l potassium chloride (2.7 mM), 1.15 g/l di-sodium hydrogen phosphate (4.3 mM), 0.2 g/l potassium dihydrogen phosphate (1.47 mM), adjusted to pH 7.4). This wash was aspirated and replaced with 2 ml

ice-cold lysis buffer {10 mM TrisHCl (pH 8.0 at 4°C)}, 130 mM NaCl, 1% Triton X-100, 10 mM NaF, 10 mM sodium phosphate, 10 mM sodium pyrophosphate). The adherent cells were scraped off using a sterile plastic cell scraper and gently transferred to a pre-cooled falcon tube. The tube was constantly agitated on a tube spinner at 4°C for 30 mins to allow full lysis to occur. The cell suspension was then centrifuged at 4°C, 15000 x *g* for 20 mins in a pre-chilled rotor. The supernatant was then transferred to a fresh tube on ice and the pellet was discarded.

The protein concentration of each cell lysate was determined by Bradford assay and the rest of each lysate was diluted with 10 µl 2X SDS sample buffer {120 mM TrisHCl (pH 6.8 at 4°C), 20 mM EDTA, 4% SDS, 0.06% Bromophenol Blue, 20% glycerol and 0.4% β-mercaptoethanol}. The cell lysate was then boiled at 100°C for 5 mins in order to denature and using the protein concentrations determined by Bradford assay, 30 µg of protein was loaded onto a SDS-PAGE gel and run as described in section 2.2.10.1.

The membrane was left overnight at 4°C in BSA blocking solution and then the membrane was incubated in primary antibody. The membrane was washed and treated with secondary antibody as described in section 2.2.10.2. Signal development was achieved by enhanced chemiluminiscence.

2.2.9 Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis Analysis (SDS-PAGE)

SDS-PAGE in Tris-glycine buffer is a technique for separating complex protein mixtures under reducing conditions. This technique was used to separate proteins within cell fractions and protein fractions generated at various stages and to estimate protein size through comparison with known protein size markers. Samples were prepared by taking 42 µl of protein fraction, adding 6 µl of 6x SDS loading buffer {360 mM TrisHCl (pH 6.8) at 4°C, 12% SDS, 60% (w/v) glycerol, 0.06% Bromophenol Blue} and 2 µl of 1M DTT. The samples along

with a 10 µl per gel aliquot of NEB broad range pre-stained protein marker were then denatured for 10 mins at 95°C and micro-centrifuged at 13000 rpm for 90 secs to separate any particulates. The supernatant (12 µl) was loaded into the pre-prepared SDS-PAGE gel wells. The gel was prepared using the component quantities described in Tables 2.5 and 2.6. The gel was submerged in 1X Tris-glycine tank buffer {5X =Tris 7.55 g, Glycine 47 g, and 10% SDS 25 ml (for 0.5 L)}

Polyacrylamide gel electrophoresis (SDS-PAGE) was used to analyse the following samples; pre-induction, post-induction obtained during protein expression, fusion protein distribution during extraction and fusion protein purity. The presence of SDS in the sample and running buffers was to ensure that proteins in the sample have the same charge: mass ratio so that the migration rate of the individual proteins in the sample will be proportional to their size. The gel was run at 30mA/70V until it passed the junction between the stacking and resolving gel. The voltage was then increased to 35mA/100V to accelerate separation. At the end of the gel run, the gel was carefully peeled off the glass plate and placed into a square Petri dish containing Coomassie InstantBlue (Expedeon). The protein bands were stained for 20 mins, the Instant Blue was discarded, and the gel washed with destaining buffer (300 ml methanol, 100 ml glacial acetic acid, 60 ml DDW), before being transferred to 2% glycerol and dried.

2.2.10 Western Blot/ Immunoblotting Analysis

This technique is used to identify the presence of and to quantify fusion-tagged protein. Native or SDS gel electrophoresis was used to separate protein samples. The separated proteins were then transferred from the gel onto a nitrocellulose membrane (Hybond-C, Amersham Biosciences) by passing current through the gel/membrane sandwich (Electro blotting), which was composed of 2x sponge pads and 2x sheets of Whatman 3MM filter paper. The components were pre-soaked in transfer buffer (2.42 g Tris, 11.52 g glycine, 160 ml methanol made up to 800 ml with DDW) before assembling the sandwich. The gel/membrane cassette was assembled in the following order; cathode, cassette, sponge, filter paper, polyacrylamide gel, nitrocellulose membrane, filter paper, sponge, cassette and anode. An ice block was added to the tank and an appropriate volume of transfer buffer to completely cover the cassette. Electro blotting was performed at 150 mA/100V for 90 mins. After transferring the protein onto the nitrocellulose membrane, the membrane was placed into a clean dish and blocked with BSA in Tris-saline (TS) buffer {10 mM TrisHCl (pH 7.0 at 4°C), 150 mM NaCl and 10 mg/ml BSA} for 45 mins. The membrane was then probed overnight with the desired primary antibody (anti-His mouse, anti-FLAG rabbit) diluted 1/2500 in blocking solution {TS buffer plus 10 mg/ml BSA plus 0.05% (w/v) Igepal}. The membrane was then washed with TS buffer for 10 mins then washed with 0.05% (w/v) Igepal in TS buffer, followed by TS buffer for 10 mins. After the washes, the membrane was subsequently probed with an anti-mouse IgG linked horseradish peroxidase, diluted1/2500 in fresh blocking solution plus 0.05% (w/v) Igepal for 2 hrs. The membrane was then washed as before. Immunodetection was performed by standard technique using chromogenic substrate chemiluminiscence. Protein bands on the membrane were highlighted by adding diaminobenzidine (DAB) prepared by dissolving one DAB tablet and one urea H2O2 tablet in 20 ml ddH2O). Upon the appearance of the bands, the staining solution was discarded, the membrane washed 3 times with DDW and stored wrapped in cling film between filter paper once dry.

2.2.11 Native PAGE Analysis

Native gel electrophoresis was used to observe the purity and migration of protein constructs in monomeric and dimeric isoforms. The samples were prepared in the same way as an SDS polyacrylamide gel, minus the SDS in the gel and loading buffer, boiling with the samples. The gel was run at pH 8.8 in order to aid the proteins migration towards the anode. Each sample (12 µl) was loaded with 4 µl of 3x native loading buffer and BSA (66 kDa) was run as a size marker. The gel was run at 30 mA (70V) until it passed the junction between the stacking and resolving gel. The voltage was then increased to 35 mA (100 V) to accelerate separation. The gel was then treated with Coomassie Instant Blue in the same manner as a SDS gel.

2.2.12 Agarose Gel

Agarose gel electrophoresis was used to identify quantity and quality of DNA preparations.
To prepare the agarose gel, 0.48 g of agarose was added to 1.2 ml 50 x TAE {242 g Tris (base), 57.1 ml glacial acetic acid and 100 ml 0.5 M EDTA} and 58.8 ml DDW. If the DNA was to be extracted from the gel after separation then boric acid was used instead of acetic acid. The solution was heated by microwaves to boiling point and mixed until all agarose had dissolved. When the solution was cooled 6 µl of ethidium bromide was added and the gel poured into the gel holder (with taped ends) and allowed to set for 30 mins. Either 1-2 µl of DNA sample or 2 µl of 1kb DNA marker ladder was added to 2 µl of 6x agarose gel loading buffer (3 ml glycerol, 25 mg bromophenol blue, 7 ml sterile DDW) and made up to 12 µl with sterile DDW. The set agarose gel was placed in the gel running apparatus, which was filled with 1X TAE buffer until the gel was submerged and the wells flooded. The DNA samples were then loaded into the wells at the negative terminal as the DNA migrates towards the positive terminal when current is applied. The gel was run at 33 mA (60V) until the ladder

had migrated the length of the gel, at which point the gel was removed and viewed under UV light.

2.2.13 Bradford Protein Assay

The Bradford protein binding assay is based on the observation that the absorbance of an acidic solution of Coomassie blue shifts from 465nm to 595nm when bound to protein. By using a standard curve of known protein concentrations, unknown protein concentrations can be determined under standard conditions.

The Bio-Rad dye reagent was supplied as a 5X concentrate and so it was first diluted to 1X with DDW and then 0.45 µm filtered. A 1.4 mg/ml bovine serum albumin (BSA) in double distilled water (DDW) solution was prepared and a serial dilution performed in preparation of the standards.

Each standard (20 µl) and 20 µl of DDW (sample buffer blank) were added to cuvettes along with 1 ml of the diluted/filtered dye reagent and gently mixed by inversion with a parafilm cover. After 10 mins the OD_{595} versus reagent blank was read (Biochrom libra S22) and a line graph of OD_{595} versus standard concentration plotted. A trend line of best fit was obtained (with an R^2 value greater than 0.95 as this ensures an accurate serial dilution has been performed) and the equation of the line used to determine unknown concentration y values (OD_{595}).

3 Production and Purification of STAT3β$_{TC}$ Proteins.

3.1 Introduction

In order to develop *in vitro* assays targeting the STAT3 pathway, STAT3 first had to be expressed recombinantly and then purified. The cloning of STAT3β$_{TC}$ into a bacterial expression vector was carried out previously by our group and so this vector was re-characterised and transformed into BL21 Rosetta and BL21 TKB1 *E. coli* strains. The results of this process will be shown and discussed in this chapter. After expression, the extracted protein was purified using a combination of affinity and ion exchange chromatography and the sample analysed using SDS PAGE gels and Western blots using poly-histidine and K15 (anti-STAT) antibodies to confirm our protein of interest.

3.2 Results and Discussion

3.2.1 Sequencing of the His-STAT3β$_{TC}$ Bacterial Expression Vector

The sequencing was performed by Eurofins MWG using a T7 primer, and the results are summarized below in Figures 3.16 and 3.17 which show the forward and reverse cloning interfaces (See appendices for the full sequence).

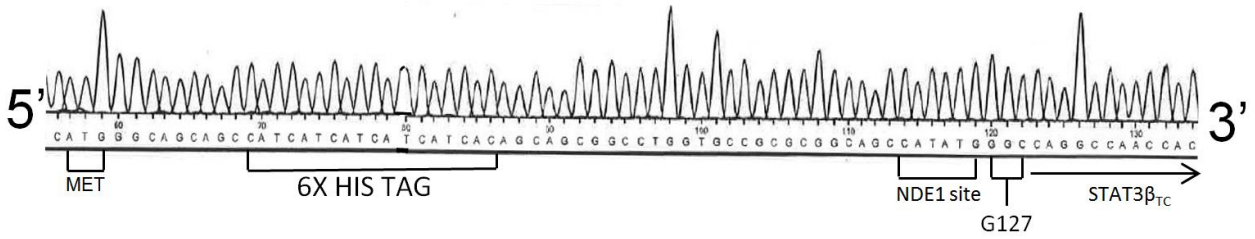


Figure 3.9 The T7 promoter 5'-3' *N*-terminal interface between the pET expression vector and the HIS-STAT3β$_{TC}$ sequence.

The translation initiating methionine codon is shown along with the *N*-terminal 6X histidine-tag. The six-base Nde1 restriction site (CATATG) indicates where the STAT3β$_{TC}$ gene was cloned in, and G127 indicates the first glycine residue of the STAT3 construct.

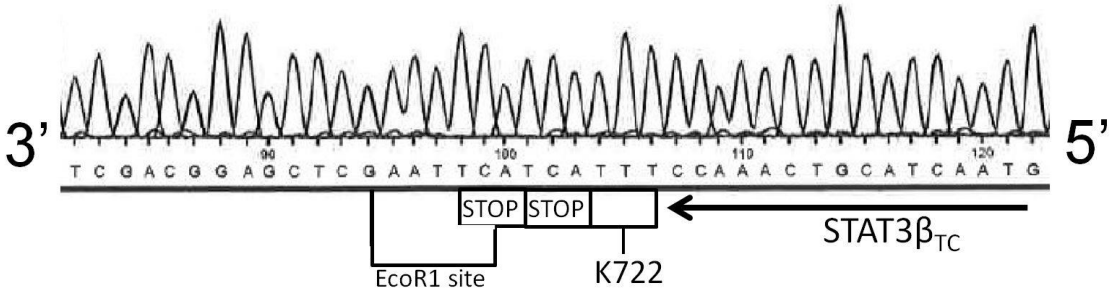


Figure 3.10 The T7 Terminator 3'-5' C-terminal interface between the pET expression vector and the HIS-STAT3β$_{TC}$ sequence.

The C-terminal end of the STAT3β_{TC} gene is shown ending with a lysine residue (K722). Translation is terminated by the two STOP codons and the EcoR1 site (GAATTC), which indicate where the STAT3β_{TC} gene was cloned into the expression vector.

The sequencing results indicate that the STAT3β_{TC} gene insert was successfully cloned into the pET28 vector, with a 6X HIS tag on the *N*-terminus. The data also confirm that a single STAT3β_{TC} gene copy is present in the correct orientation and reading frame. The HIS-STAT3β_{TC} construct consists of 610 amino acids i.e.21 pET 28c vector bases, including the 6X histidine tag, followed by 589 amino acids of STAT3β (127-722 inclusive), and has a theoretical molecular weight of 70.4 kDa and an isoelectric point of pH 7.58.

3.2.2 Expression of Un-Phosphorylated HIS-STAT3β_{TC}

The HIS-STAT3β_{TC} bacterial expression vector was transformed into competent BL21 DE3 Rosetta *E. coli* cells. Expression was induced with 1 mM IPTG, as described in Section 2.2.2.1, leading to over-expression of the un-phosphorylated form of the HIS-STAT3β_{TC} protein.

The HIS-STAT3β_{TC} consistently appeared just below the 62kDA band of the pre-stained marker despite having a theoretical molecular weight of 70.4kDa. In addition, the mass calculated from mass spectrometry resulted in a mass of 68 kDa (fig 3.20) and it is this mass that is thought to be more accurate. The discrepancy could be due to inaccuracies in the pre-stained markers migration and also the protein may bind less SDS than expected due to incomplete unfolding and hence appear to have less mass. The protein was fully sequenced, as shown in appendix 9.1.

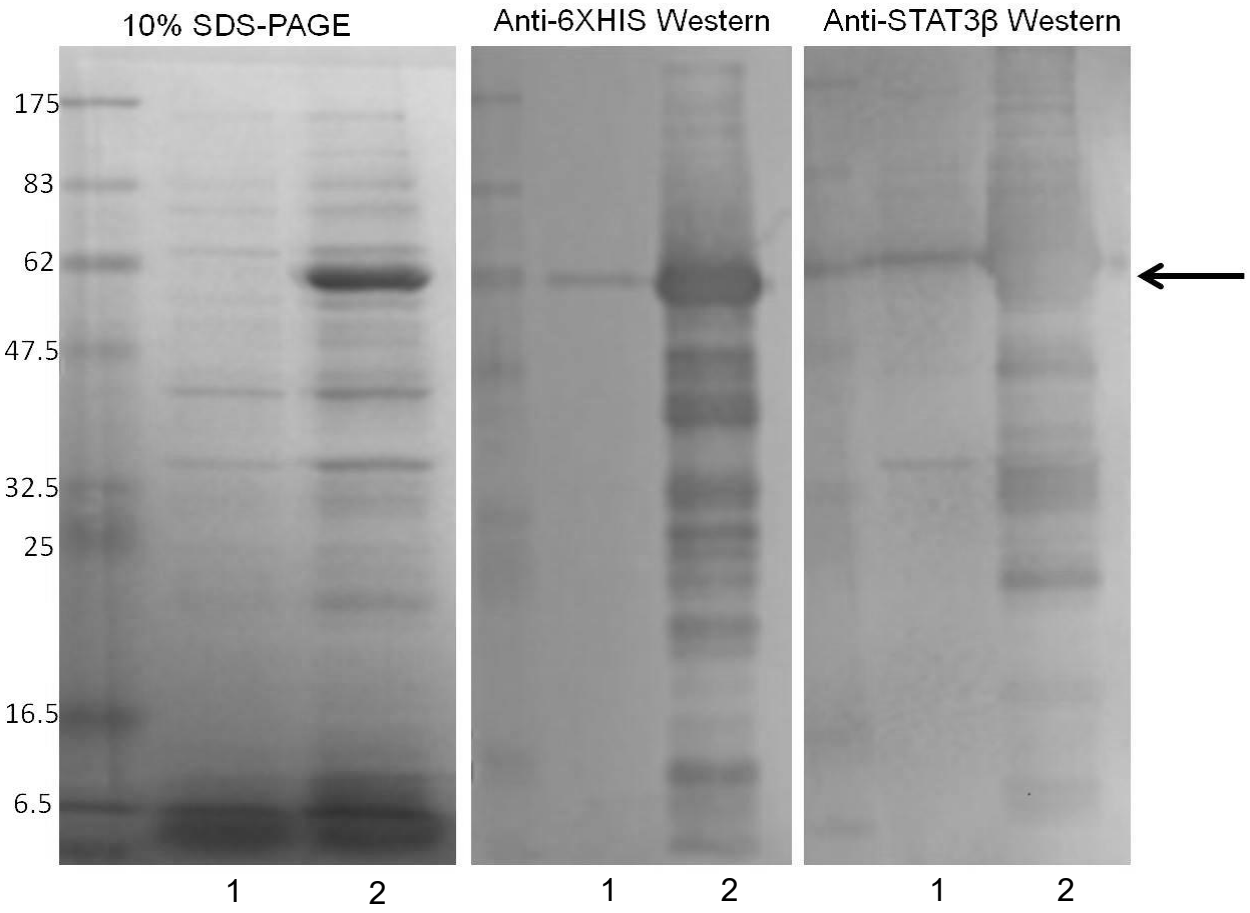


Figure 3.11 Pre and post (lanes 1 and 2 respectively) IPTG induction samples of un-phosphorylated HIS-STAT3β$_{TC}$ production visualised with Coomassie blue stain for SDS gel and DAB solution for Western blots, as detailed in Sections 2.2.9/10.

Figure 3.18 shows a 10% SDS-PAGE gel along with Anti-6XHIS and Anti-STAT3β$_{TC}$ (AA 626-640) K15 Western blots. Lane 1 (left to right) is a broad range pre-stained protein marker (6-175kDa), lane 2 is a pre-IPTG induction sample, and lane 3 is a post-IPTG induction sample. The over-expressed HIS-STAT3β$_{TC}$ protein is indicated by an arrow.

A large increase of HIS-STAT3β$_{TC}$ is seen at the correct approximate molecular weight (i.e.70 kDa) when pre-IPTG induction and post-IPTG induction are compared in the 10% SDS-PAGE gel. The Western blots confirm that the induced protein has a 6X HIS tag and that the expressed protein is of the STAT3 family, indicating a successful over-expression of HIS-STAT3β$_{TC.}$

3.2.3 Purification of Un-Phosphorylated HIS-STAT3β$_{TC}$

Extraction of HIS-STAT3β$_{TC}$ and subsequent purification was achieved through ion-exchange chromatography, which resulted in pure HIS-STAT3β fractions when visualized by SDS-PAGE and confirmed with Anti-6XHIS Western blot (Figure 3.19).

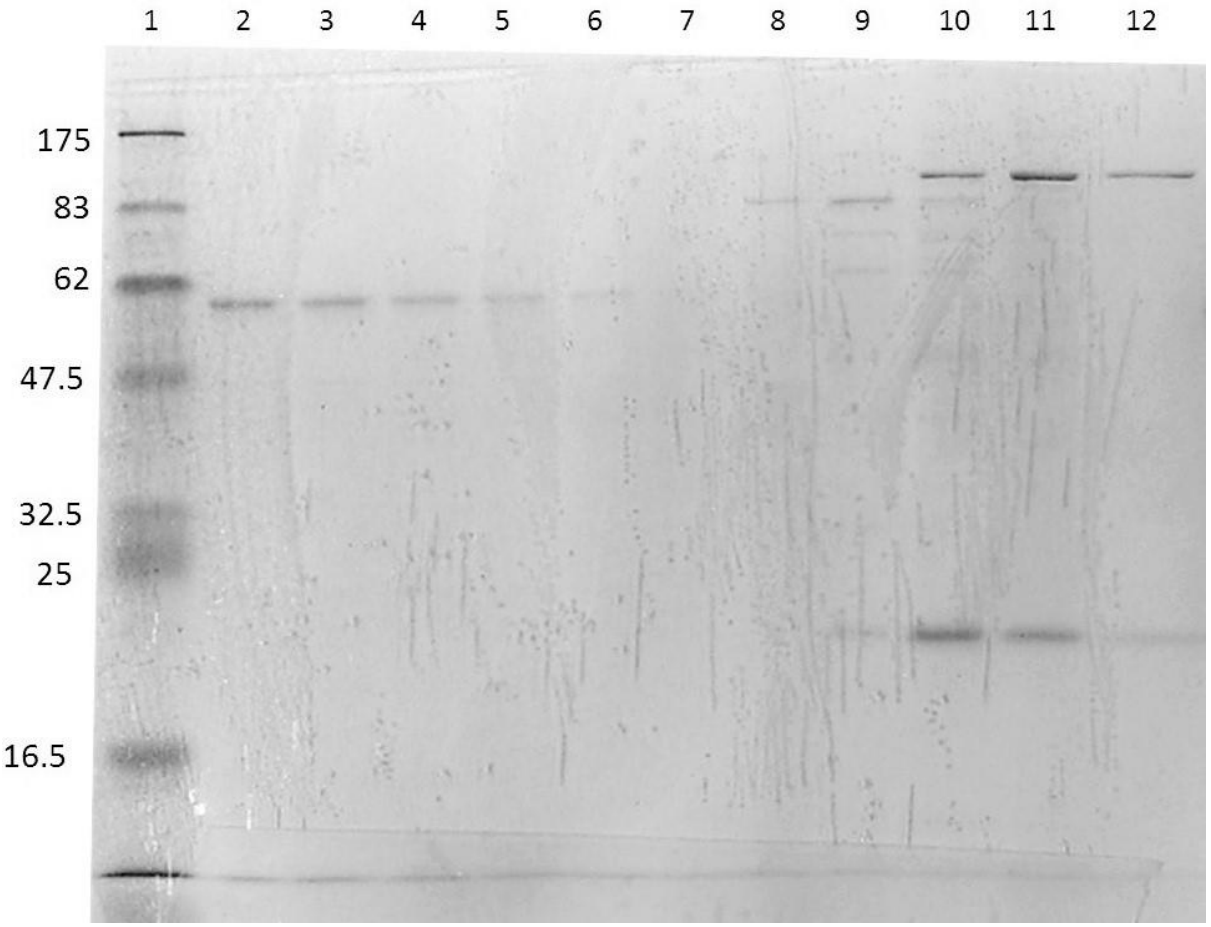


Figure 3.12 10% SDS PAGE gel showing ion exchange chromatography purification of un-phosphorylated HIS-STAT3β$_{TC}$. Lane 1: Broad range pre-stained protein marker (6-175 kDa); Lane 2: F1 1 ml 0.1 M NaCl; Lane 3: F2 1 ml 0.1 M NaCl; Lane 4: F3 1 ml 0.1 M NaCl; Lane 5: F4 1 ml 0.1 M NaCl; Lane 6: F5 1 ml 0.1 M NaCl; Lane 7: F6 1 ml 0.1 M NaCl; Lane 8: 10 ml 0.2 M NaCl; Lane 9: 10 ml 0.3 M NaCl; Lane 10: 10 ml 0.4 M NaCl; Lane 11: 10 ml 0.5 M NaCl; Lane 12: 10 ml 1 M NaCl; visualised with Coomassie blue stain, as detailed in Section 2.2.9

The purified HIS-STAT3β_{TC} was eluted by 0.1M NaCl in 1 ml fractions and fractions 4-6 reserved for protein concentration determination.

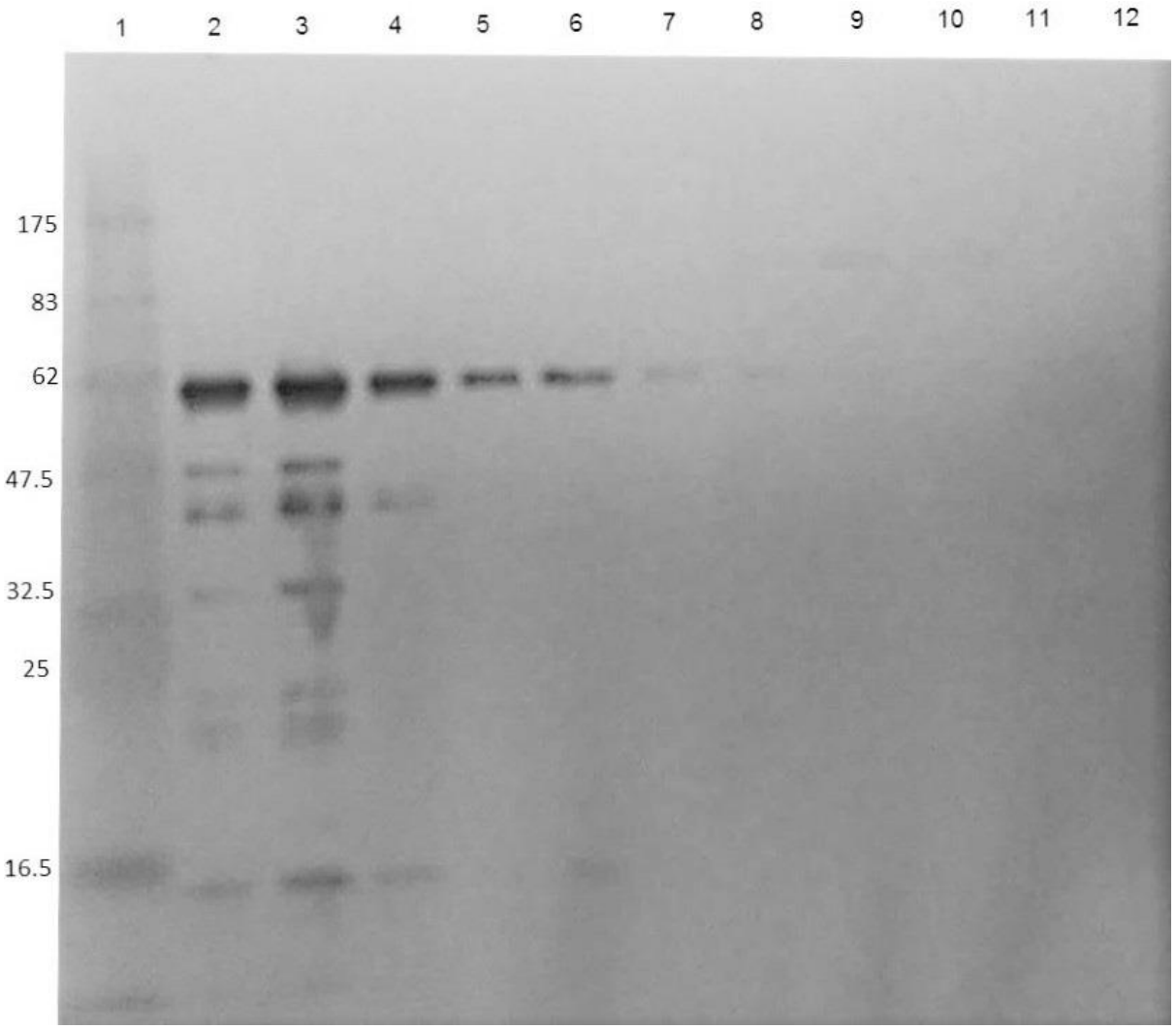


Figure 3.13 Anti-6XHIS Western showing ion exchange chromatography purification of un-phosphorylated HIS-STAT3β_{TC}. Lane 1: Broad range pre-stained protein marker (6-175 kDa), Lane 2: F1 1 ml 0.1 M NaCl, Lane 3: F2 1 ml 0.1 M NaCl, Lane 4: F3 1 ml 0.1 M NaCl, Lane 5: F4 1 ml 0.1 M NaCl, Lane 6: F5 1 ml 0.1 M NaCl, Lane 7: F6 1 ml 0.1 M NaCl, Lane 8: 10 ml 0.2 M NaCl, Lane 9: 10 ml 0.3 M NaCl, Lane 10: 10 ml 0.4 M NaCl, Lane 11: 10 ml 0.5 M NaCl, Lane 12: 10 ml 1 M NaCl, visualised with Coomassie blue stain, as described in Section 2.2.10.

The lower molecular weight bands seen in fractions 2 and 3 could be due to either protein degradation or nonspecific HIS antibody binding. Therefore, purer fractions (5-6) with less concentrated protein of interest were used to carry out subsequent assays.

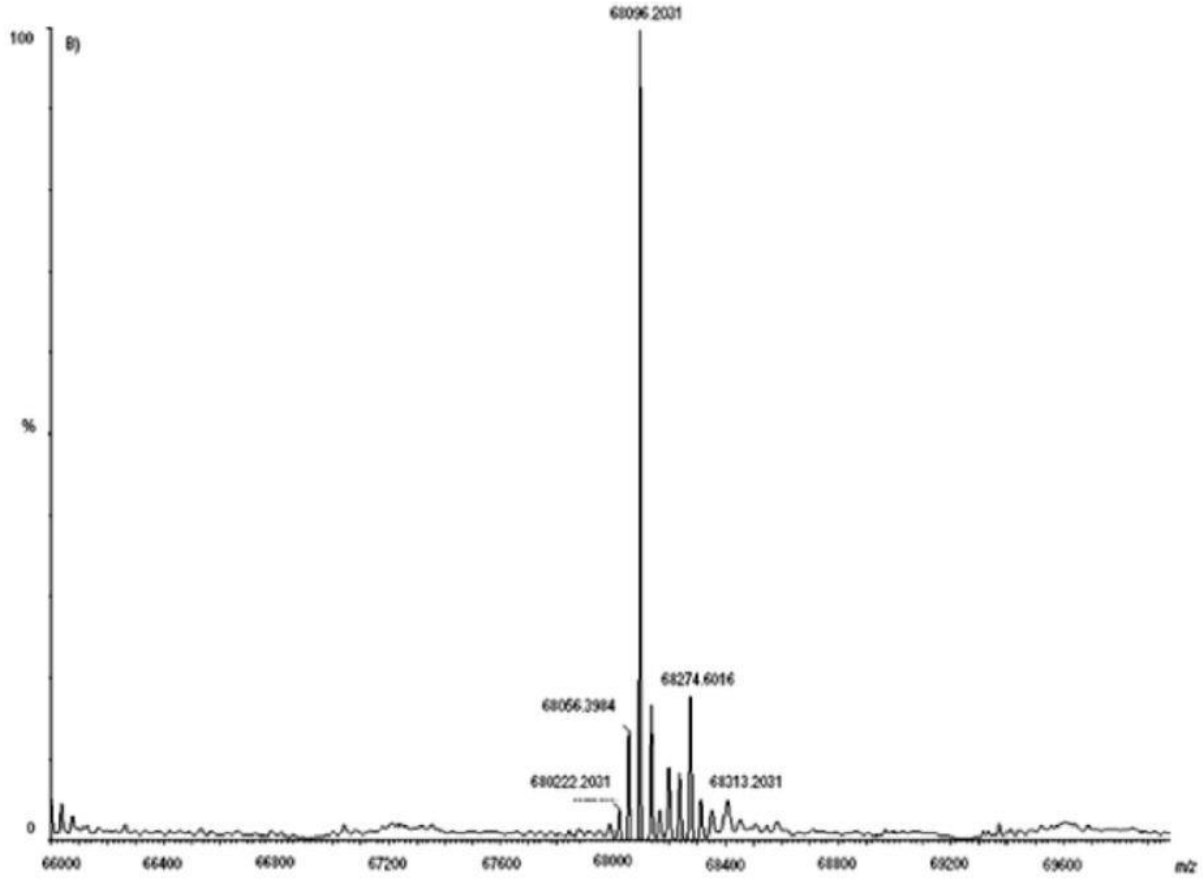


Figure 3.14 De-convoluted ESI mass spectrum of purified unphosphorylated STAT3β$_{TC}$ protein acquired using Q-TOF (Data provided by Dr Sibylle Heidelberger).

ESI mass spectra usually contain same species molecules with different charge states. Deconvolution was carried out on this spectrum so that the multiple-charged species were recalculated into a singly charged form and grouped together according to their peak width. The spectrum obtained was consistent with the calculated molecular mass of 68,098 Da (error = 30 ppm) for unphosphorylated STAT3β$_{TC}$**.** This information was used in conjunction with sequencing data shown in appendices 9.1 and 9.2 to confirm the identity of STAT3β$_{TC.}$

3.2.4 Expression of Phosphorylated HIS-STAT3β$_{TC}$

The HIS-STAT3β$_{TC}$ pET 28c vector was transformed into competent TKB1 BL21-derived *E. coli* cells. Expression was induced with 1 mM IPTG (as described in Section 2.2.3.2).The bacteria were pelleted and kinasing media added resulting in the expressed kinase phosphorylating tyrosine residue 705 in the SH-2 domain of the over-expressed HIS-STAT3β$_{TC}$ protein. This tyrosine residue is the region of the SH-2 domain implicated in the homodimerisation of STAT3β$_{TC}$.

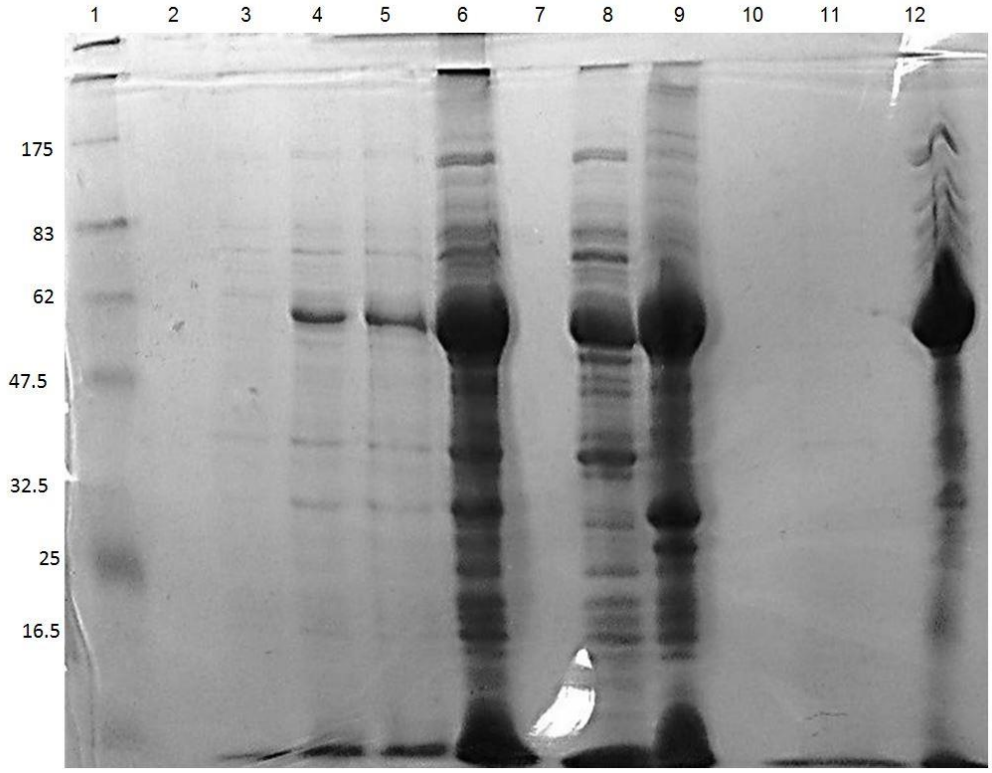


Figure 3.15 10% SDS-PAGE gel (2.2.9) of the phosphorylated HIS-STAT3β$_{TC}$ (pSTAT) expression and extraction steps (2.2.2.2) Lane 1: Broad range pre-stained protein marker (6-175kDa), Lane 2: Empty, Lane 3: Pre-IPTG induction, Lane 4: Post-IPTG induction, Lane 5: Post kinasing step, Lane 6: Post sonication, Lane 7: Empty, Lane 8: Post sonication supernatant, Lane 9: Post sonication pellet, Lane 10: Empty, Lane 11: Post ammonium sulphate precipitation supernatant, Lane 12: Redissolved ammonium sulphate pellet, visualised with Coomassie Blue stain.

The IPTG induced over-expression of HIS-STAT3β_{TC} is clearly seen between Lanes 3 and 4. Lane 6 shows the significant release of material from the cells after sonication. Lanes 8 and 9 show that although large amounts of HIS-STAT3β_{TC} remain in the insoluble pellet, there is a significant quantity in the soluble supernatant, and this step acts as an initial purification step. Lanes 11 and 12 show that the ammonium sulphate precipitation worked efficiently with no visible protein left in the supernatant, and that the re-dissolved pellet had fractionated ready for further purification.

3.2.5 Ion Exchange Chromatography Purification of HIS-STAT3β$_{TC}$

Purification of phosphorylated HIS-STAT3β$_{TC}$ was achieved through ion-exchange, His-trap affinity and size exclusion chromatographic methods.

Figure 3.23 shows a typical gel resulting from ion exchange purification which indicated relatively high purity. However, the recovery yield was relatively poor in relation to the starting material.

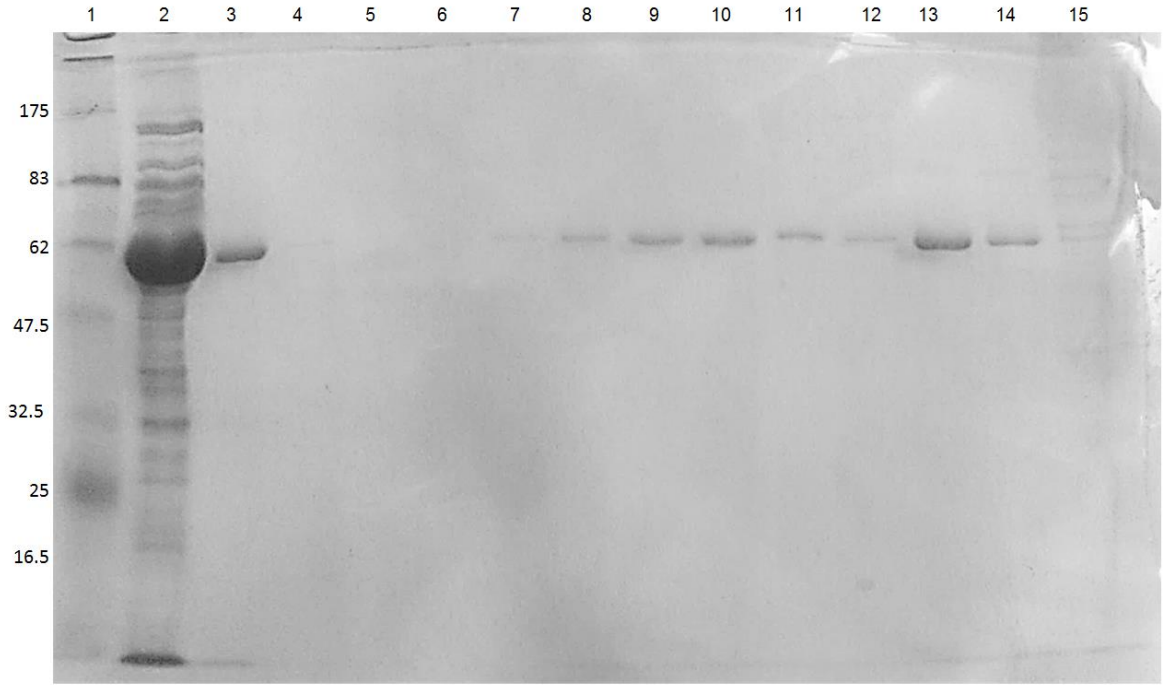


Figure 3.16 10% SDS-PAGE gel of phosphorylated HIS-STAT3β$_{TC}$ (pSTAT) purification using ion exchange chromatography. Lane 1: Broad range pre-stained protein marker (6-175kDa), Lane 2: re-dissolved ammonium sulphate pellet, Lane 3: Dissolved HIS-STAT3β$_{TC}$ column flow through, Lane 4: Column wash flow through, Lane 5: F1 1 ml 0.1 M NaCl, Lane 6: F2 1 ml 0.1 M NaCl, Lane 7: F3 1 ml 0.1 M NaCl, Lane 8: F4 1 ml 0.1 M NaCl, Lane 9: F5 1 ml 0.1 M NaCl, Lane 10: F6 1 ml 0.1M NaCl, Lane 11: F7 1 ml 0.1 M NaCl, Lane 12: F8 1 ml 0.1 M NaCl, Lane 13: 10 ml 0.2 M NaCl elution, Lane 14: 10 ml 0.3 M NaCl elution, Lane 15: 10 ml 0.4 M NaCl elution, visualised with Coomassie Blue stain.

3.2.6 His-Trap Affinity Chromatography Purification of HIS-STAT3β$_{TC}$

His trap affinity chromatography followed by size exclusion purification was used as a purification strategy in an attempt to improve yield.

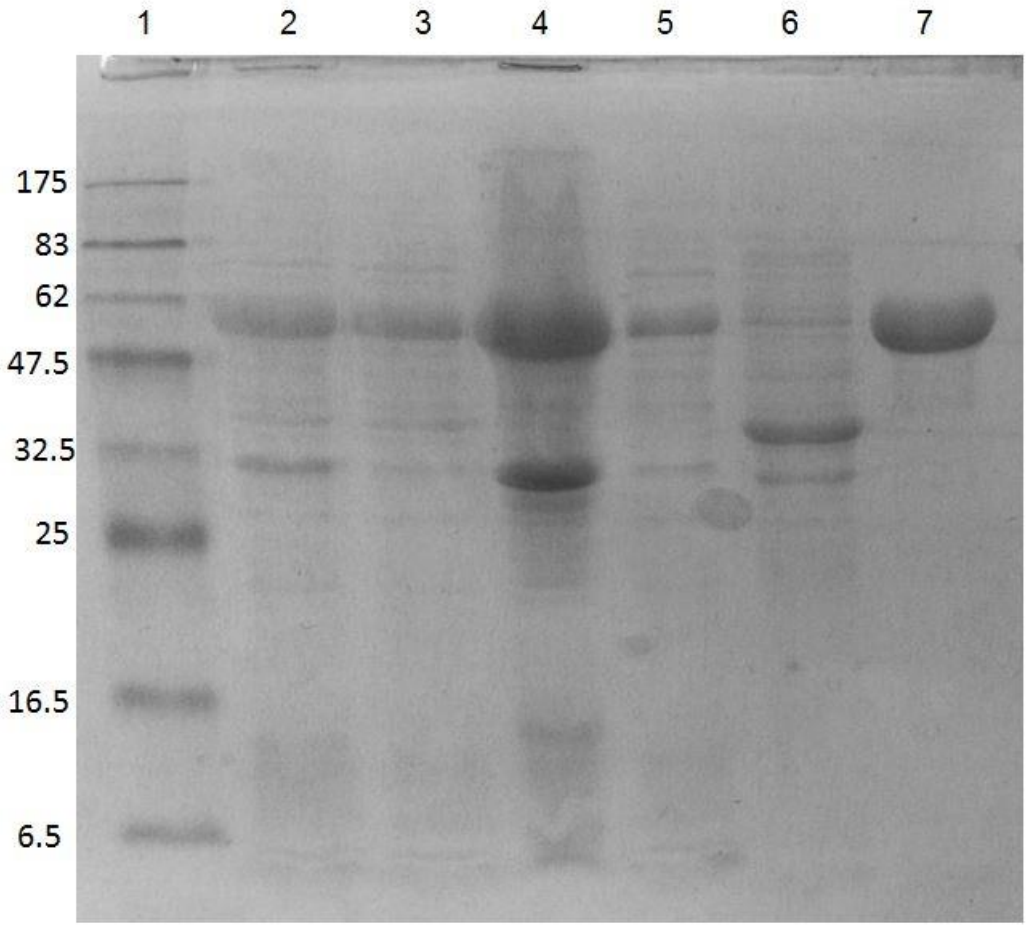


Figure 3.17 10% SDS-PAGE gel of phosphorylated HIS-STAT3β$_{TC}$ (pSTAT) extracted and purified by His-Trap affinity chromatography. Lane 1: Broad range pre-stained protein marker (6-175kDa), Lane 2: Post sonication lysate, Lane 3: Post sonication supernatant, Lane 4: Post sonication pellet, Lane 5: His Trap loading flow through, Lane 6: His Trap wash, Lane 7: HIS-STAT3β$_{TC}$ elution, visualised with Coomassie Blue stain, as described in Section 2.2.9.

The gel demonstrates that the extraction process leaves a large proportion of HIS-STAT3β_{TC} in the cell pellet, and this could be due to the presence of excess protein forming hard-to-solubilise inclusion bodies (174). However, by passing the supernatant through the His-Trap, the protein was concentrated and eluted in a small elution volume. The subsequent His-Trap wash showed that many non-specific impurities had been removed.

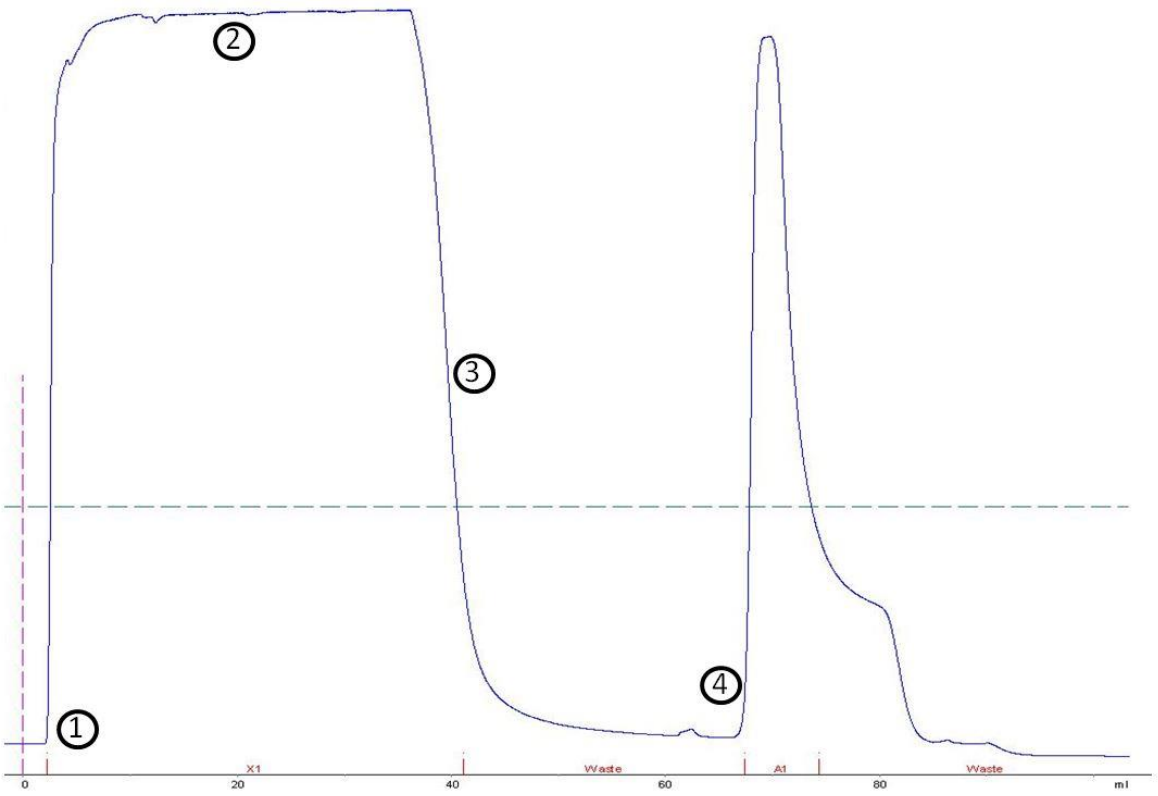


Figure 3.18 His-Trap affinity purification chromatogram of phosphorylated HIS-STAT3β_{TC}.

In Figure 3.25, the X axis represents the volume (ml) of solvent to pass through the column, and the Y axis represents the UV absorbance wavelength with the absorption intensity indicating the amount of protein in the sample. Point 1 In the figure indicates where the HIS-STAT3β_{TC} sample was injected onto the column, with the absorbance rapidly rising from the buffer only baseline. Point 2 shows the sample continuing to be loaded, and reaches a plateau at the absorbance of the impurities eluting from the column. Point 3 represents the column wash during which non-bound proteins are washed off the column with equilibration

buffer until the UV absorbance decreases to a steady baseline. Finally point 4 represents the start of the protein elution using a solution of 500 µM imidazole. The protein was eluted rapidly as one large peak, which was collected as fraction A1. Collection was stopped before the shoulder peak was eluted to maintain purity.

3.2.7 Size Exclusion Purification of HIS-STAT3β$_{TC}$

Size exclusion chromatography (SEC) was carried out with a S12 column (See figure 2.26), as described in Section 2.2.2.3, to purify the protein further.

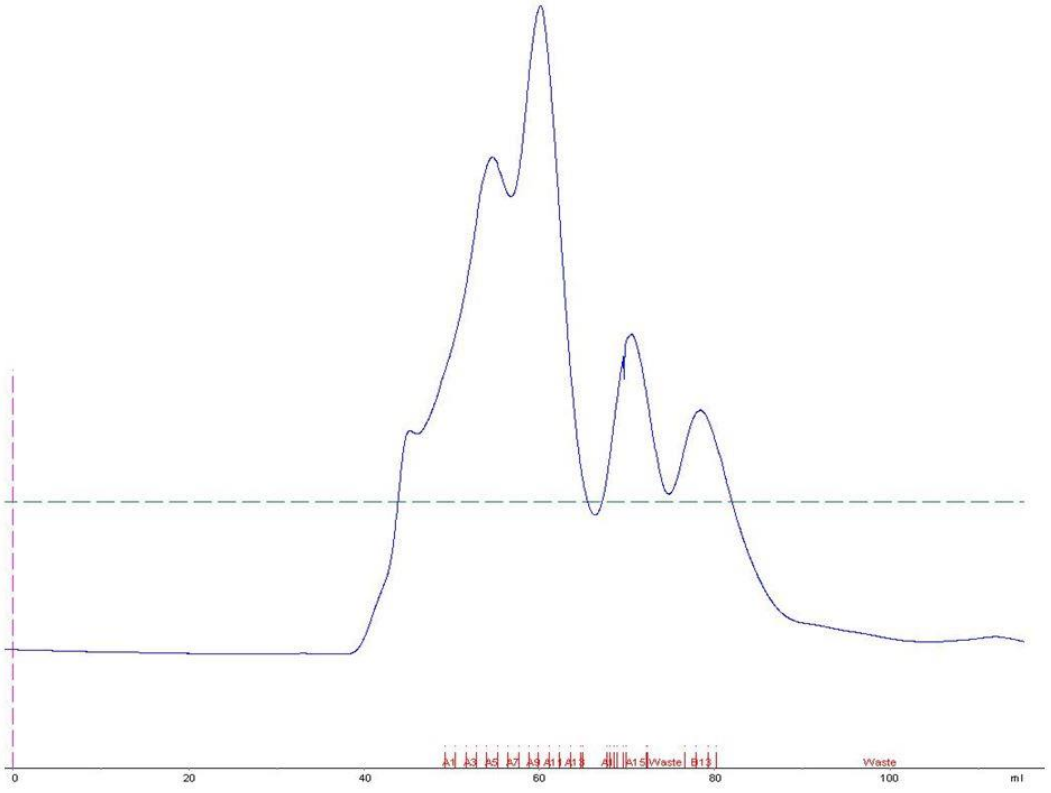


Figure 3.19 Gel filtration (S12) chromatogram showing the further purification of phosphorylated HIS-STAT3β$_{TC}$ by size exclusion chromatography.

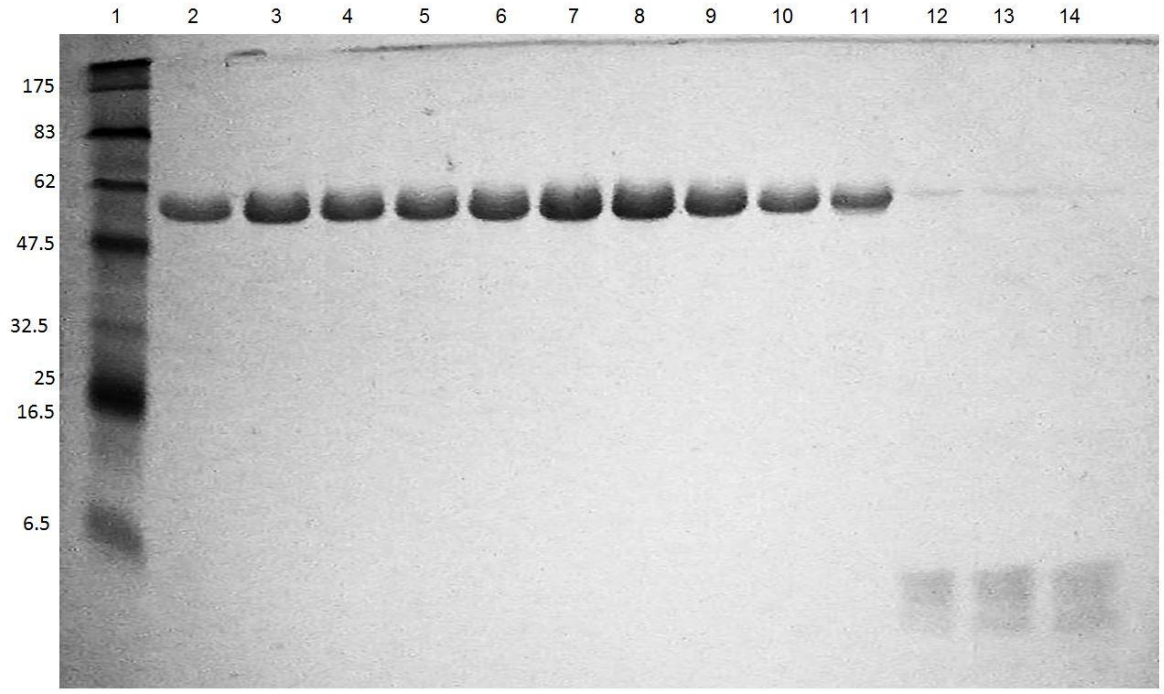


Figure 3.20 10% SDS-PAGE gel of phosphorylated HIS-STAT3β$_{TC}$ (pSTAT) further purified by size exclusion chromatography. Lane 1: Broad range pre-stained protein marker (6-175kDa), Lanes 2 through 11: 1 ml gel filtration fractions A4-A15, Lanes 12 through 14: 1 ml gel filtration fractions B12 through B14, visualised with Coomassie Blue stain, as described in Section 2.2.9.

Once the extracted protein was purified and analysed the protein fractions with estimated highest purity and yield were selected to determine protein concentration by Bradford assay. The following chapter describes this process and the validation of the homogenous time-resolved FRET (HTRF) assay.

4 Development of Homogenous Time-Resolved FRET (HTRF) Assay

4.1 Introduction

This chapter describes the development and optimization of a homogenous time-resolved fluorescence (HTRF) assay the basis of which is represented in Figures 4.28 and 4.29. An HTRF assay was chosen as it is not reliant on washing and binding surfaces, and miniaturization choices are available increasing throughput as well as reducing the amounts of purified protein and costly reagents required. HTRF also limits the short-lived background fluorescence often associated with FRET experiments by introducing a time delay between the system excitation and the fluorescence measurement. Thus, the short-lived non-specific fluorescence is not measured and the stable fluorescence of the HTRF fluorophores is detected. In this way the quenching effect is also minimized.

The determination of STAT3β_{TC} concentration is described below, followed by a discussion of the initial set-up and validation of the Envision plate reader. Comparisons using biotinylated *versus* non-biotinylated and phosphorylated HIS-STAT3β_{TC} *versus* unphosphorylated HIS-STAT3β_{TC} are also described as examples of assay specificity. All recombinant STAT3 proteins were >95% purity, as estimated by gel filtration. Optimisation of the use of potassium fluoride and time-to-equilibrate are also included. Finally, evaluation of the selectivity of the assay using the positive control inhibitors STA-21 and SH-2 domain peptide mimics is also described.

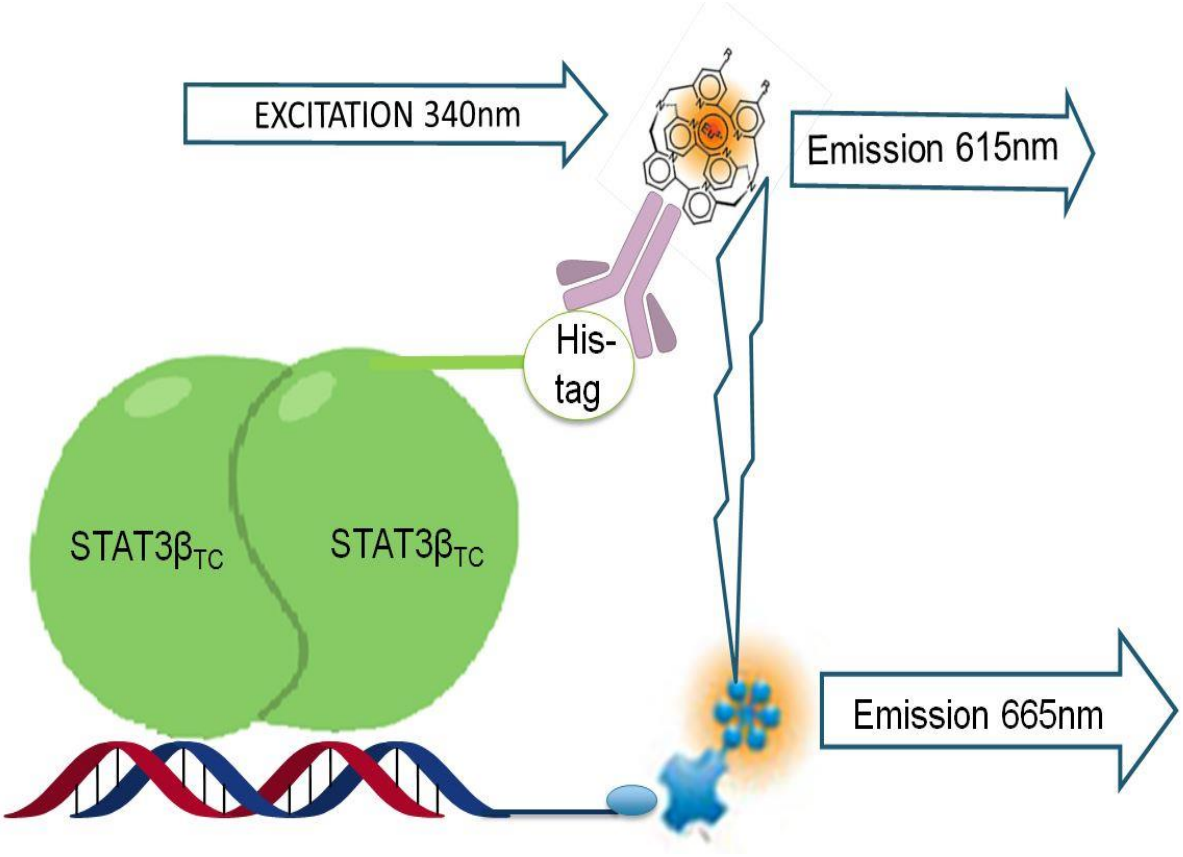


Figure 4.21 Schematic diagram of the STAT3β$_{TC}$ HTRF assay.

The HIS-tagged pSTAT3β$_{TC}$ homodimer is bound to its constitutive biotinylated DNA sequence and a streptavidin-d2 conjugate binds covalently to the biotin whereas an anti-HIS antibody conjugated with a Eu^{3+}-cryptate fluorophore binds to the HIS tagged pSTAT monomer. The Europium-cryptate ligand is excited at 340 nm, and FRET then occurs between the fluorophores with the d2 fluorophore emitting fluorescence at 665 nm. Background emission from the europium is observed at 615 nm.

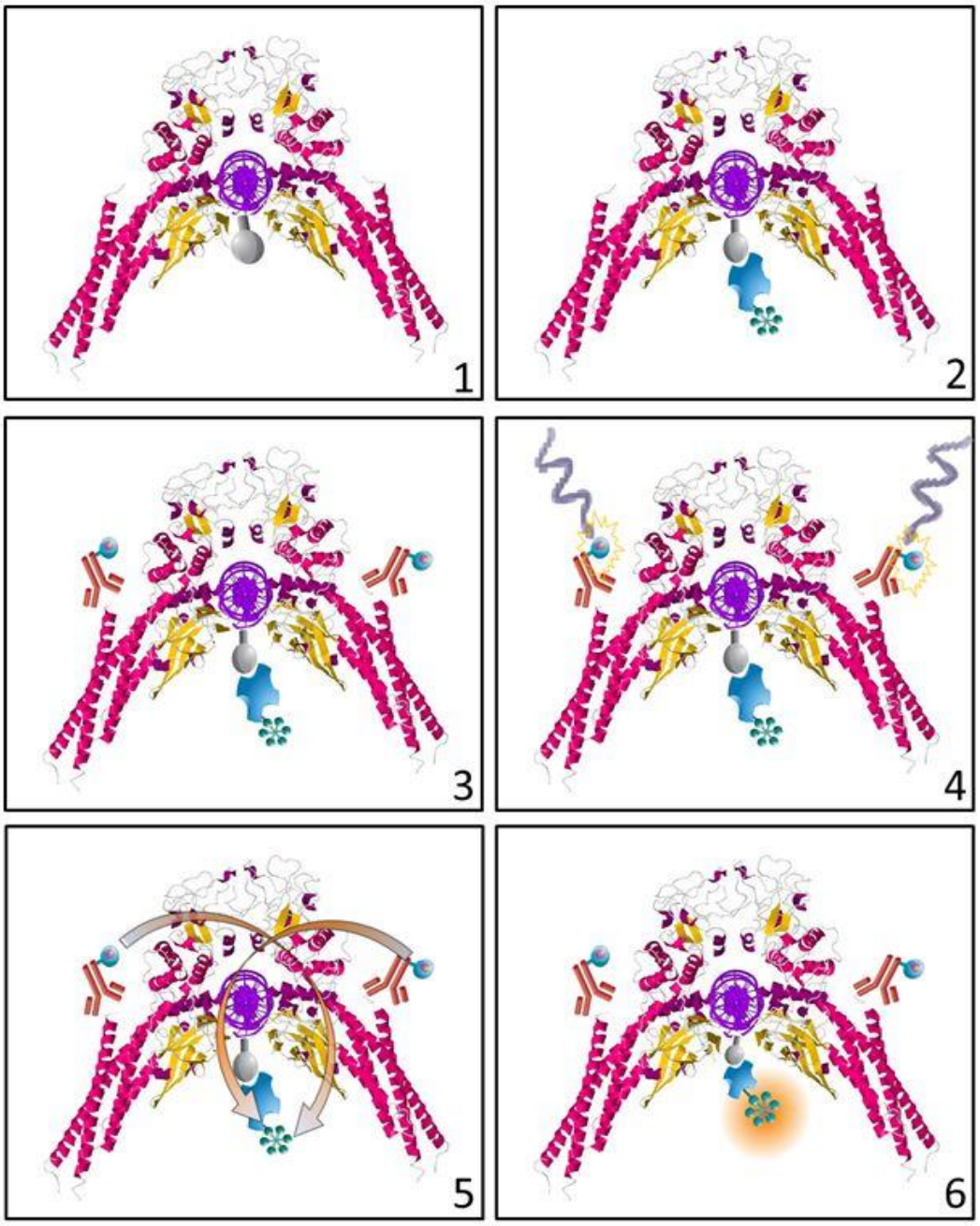


Figure 4.22 Components of the HTRF assay and sequence of addition.

1) pSTAT homodimer bound to DNA (purple). The grey construct attached to the end of the oligo represents biotin; 2) Addition of the streptavidin-d2 conjugate leads to it binding covalently to the biotin; 3) Addition of anti-HIS antibody conjugated to the Eu^{3+}-cryptate fluorophore leads to it binding to the N-terminal of each pSTAT monomer; 4) The europium cryptate ligand is excited at 320 nm; 5) FRET occurs between the fluorophores; 6) The d2 fluorophore emits fluorescence at 665 nm.

4.2 Results and Discussion

4.2.1 Concentration of STAT Protein Determined by Bradford Assay

Accurate determination of recombinant protein concentrations was achieved using Bradford Assay, as described in Section 2.2.10.5. The equation was rearranged to make "X" the subject and the unknown protein solution OD_{595} was swapped into the equation as the "Y" value to generate the unknown protein concentration in mg/ml. The mg/ml concentration was then converted to molar concentration by dividing by the molecular weight of the protein.

4.2.2 Plate Reader Initial Set-up

The Wallac EnVision 2101 plate reader was set-up as follows, with the assistance of Cisbio technical support to ensure optimal HTRF readout: Ideally a 340 nm excitation filter and 620 nm emission filter would have been used for greatest optimisation.

Condition	Setting
Top mirror	LANCE/DELFIA
Excitation filter	UV2 (TRF) 320 nm
Emission filter	APC 665 nm
2nd Emission filter	Europium 615 nm
Measurement height (mm)	6.5
Excitation light (%)	100
Delay (µs)	60
Window time (µs)	100
Time between flashes (µs)	400
Number of flashes	100

Table 4.9. Envision parameter set-up

4.2.3 Plate Reader Suitability Study

A homogenous time-resolved fluorescence (HTRF) reader control kit (Cisbio Bioassays, France) was used to assess the suitability of the Perkin Elmer Multilabel plate reader (Wallac EnVision 2101) and associated filter set, to validate their ability to integrate with HTRF technology. This kit allows the set-up of overall performance as well as management of the lifetime of the excitation source and counting rate for the reader. The kit was run on a 96 well half-volume plate in the same way that the full assay was carried out.

The test consisted of two steps, the first of which aimed to evaluate the signal/blank (S/B) ratio using a calibrated 620 nm control. If this initial test were passed then the second phase was to evaluate the equipment's detection limits.

The 620 nm control and blank buffer were added in triplicate to the plate and then read at 615 and 665 nm. The S/B value was calculated using the following formula:

$$S/B = \frac{\text{Count per second (cps) 620 nm}_{\text{control}}}{\text{Count per second (cps) 620 nm}_{\text{buffer blank}}}$$

The results are shown below in Table 4.12. The signal/blank ratio of 79 showed that there was a significant fluorescence background signal (i.e.> 40). This signal was a simple demonstration of excitation and emission detection. The second test, based on the background signal, investigated the signal's ability to be transferred to another fluorophore (FRET) emitted and detected.

	665 nm		620 nm		Norm
	Mean	CV %	Mean	CV %	
Buffer blank	263.66	7.02	111.33	14.41	
620 nm control	1610.75	3.34	8824.75	2.41	
S/B			79.26		≥ 40

Table 4.10. Envision 620 nm signal/background ratio

The control of the detection limit was tested on the same 96 well, low volume plate and introduced the two fluorophores (Eu-cryptate and d2) at calibrated levels. A standard blank (Eu-cryptate and d2 diluted in reconstitution buffer = No FRET) was compared with a low control (Eu-cryptate and d2 with low calibrator = low FRET) and a high control (Eu-cryptate and d2 with high calibrator = maximum FRET). The means and standard deviations for each sample set were calculated at both wavelengths (615 and 665 nm). The ratio (665 nm/615 nm) was calculated as follows:

$$Ratio = \frac{cps_{665\,nm}}{cps_{615\,nm}} \times 1000$$

From this ratio, ΔR and ΔF for both the low and high controls are calculated. ΔR is given by the formula:

$$\Delta R = R_{Calx} - R_{std0}$$

Where "Cal*x*" is either the low or high calibrator. ΔF (which represents the S-B/B) was obtained with the formula:

$$\Delta F = \frac{\Delta R}{R_{std0}} \times 100$$

The ΔF for the low and high calibrators at 3 and 18 hour time points are shown below in Table 4.13

Incubation time :	[x] 3 hours		[x] 18 hours	
	DF %	Norm	DF %	Norm
Low calibrator	20.71	≥ 10 %	35.70	≥ 15 %
High calibrator	447.81	≥ 500 %	1138.25	≥ 600 %

Table 4.11.ΔF values for low and high FRET calibrators

The calibrators were all well within the normal ranges, except the 3 hrs high calibrator, which at 447% was 10% below the normal 500%. This difference was not deemed overly significant with time to equilibrate over 18 hrs. The signal became nearly 100% over the expected norm. Table 4.14 shows the control figures of the individual standards in isolation. These are a secondary check in case the low and high calibrators fall well beyond the norms. They are included here to show the large emission seen at 620 nm for the Eu-cryptate when compared with the XL665 fluorophore. Also of note is the low XL665 emission signal at 665 nm demonstrating its lack of excitation at 320 nm and, therefore, its suitability as a FRET fluorophore with Europium cryptate.

	CV % ratio	Norm
Standard 0	4.31	≤ 10 %

	665 nm		620 nm	
	Mean	CV %	Mean	CV %
Buffer blank	280	6.55	115	11.37
Cryptate blank	751	2.92	4121	3.40
XL665 blank	297	2.29	111	5.52

Table 4.12. Emission readings of individual components at 320 nm excitation

4.2.4 Non-Biotinylated DNA Negative Control Study

The HTRF assay was run using two phosphorylated HIS-STAT3β$_{TC}$ concentrations (10 and 100 nM and compared with M67 double stranded DNA that was biotinylated and also un-biotinylated DNA (Figure 4.31).

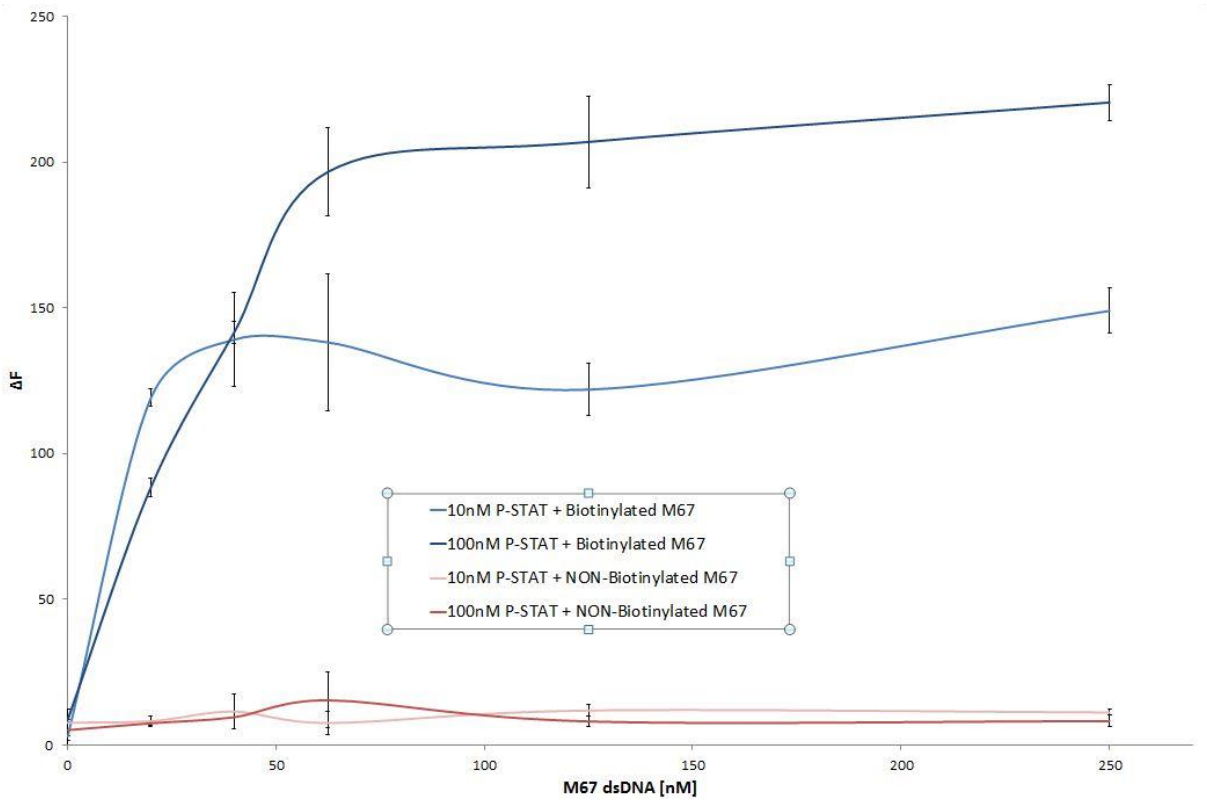


Figure 4.23. Line graph showing effect of biotinylated DNA in HTRF assay with >95% purity P-STAT3β$_{TC.}$ (HTRF assay executed as described in 2.2.5. Means derived from N=3 +/- SD)

The assay shows that over a range of DNA concentrations (20-250 nM) if the DNA used lacks biotinylation then no increase in FRET signal is seen compared with the fluorophores only control. This was the case at both pSTAT concentrations tested (10 and 100 nM). Biotin on the DNA is required for the streptavidin –d2 conjugate to bind to in order to tether the

fluorophore to the transcription complex and for significant FRET to occur between fluorophores. Significant FRET signals are seen when biotinylated DNA is used and this signal is proportional to the amount of pSTAT used, but the signal plateaus, so the signal is proportional to both the protein and DNA present and also suggesting that the FRET signal is due to the interaction between pSTAT and biotinylated M67 DNA and not a non-specific binding event.

4.2.5 Phosphorylated STAT3β$_{TC}$ is Required for Assay Activity

Recombinant phosphorylated HIS-STAT3β$_{TC}$ was compared with recombinant unphosphorylated HIS-STAT3β$_{TC}$ at four concentration points (0.1 to 100 nM). Biotinylated DNA was kept standard at two different concentrations (20 and 40 nM) (Figure 4.32)

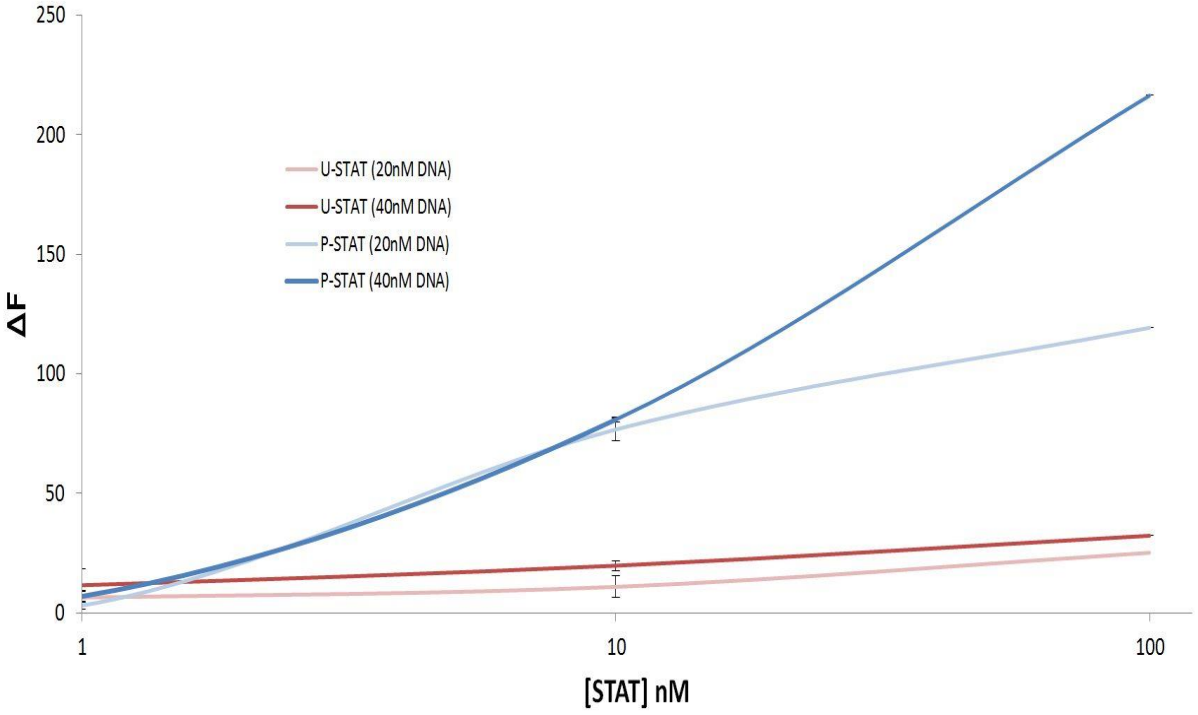


Figure 4.24 Line graph comparing phosphorylated and unphosphorylated HIS-STAT3β$_{TC}$ use in HTRF assay with consistent biotinylated DNA. (HTRF assay executed as described in 2.2.5. Means derived from N=3 +/- SD)

The FRET signal remains consistently low when using unphosphorylated HIS-STAT3β_{TC}, with a marginal increase in ΔF between 10 and 100 nM uSTAT of 15 and 13 for 20 and 40 nM DNA sample. In contrast, the increase in ΔF between 10 and 100 nM with pSTAT was 43 and 136 for 20 and 40 nM DNA respectively. This increase in signal is hypothesized as the binding ratio of 2:1 protein:DNA is achieved. The signal is dependent on a homodimer forming between two pSTAT monomers and subsequent binding to DNA and, therefore, use of uSTAT was not pursued further.

4.2.6 Reagent Time Study

The effect of allowing the assay components to equilibrate over time was observed. Components were added in the normal way and concentration (i.e. 100 nM pSTAT, 50 nM biotinylated DNA). Samples were read immediately and then every hour over a six hour period. 0.4M Potassium fluoride was added as standard just before FRET reading (Figure 4.33).

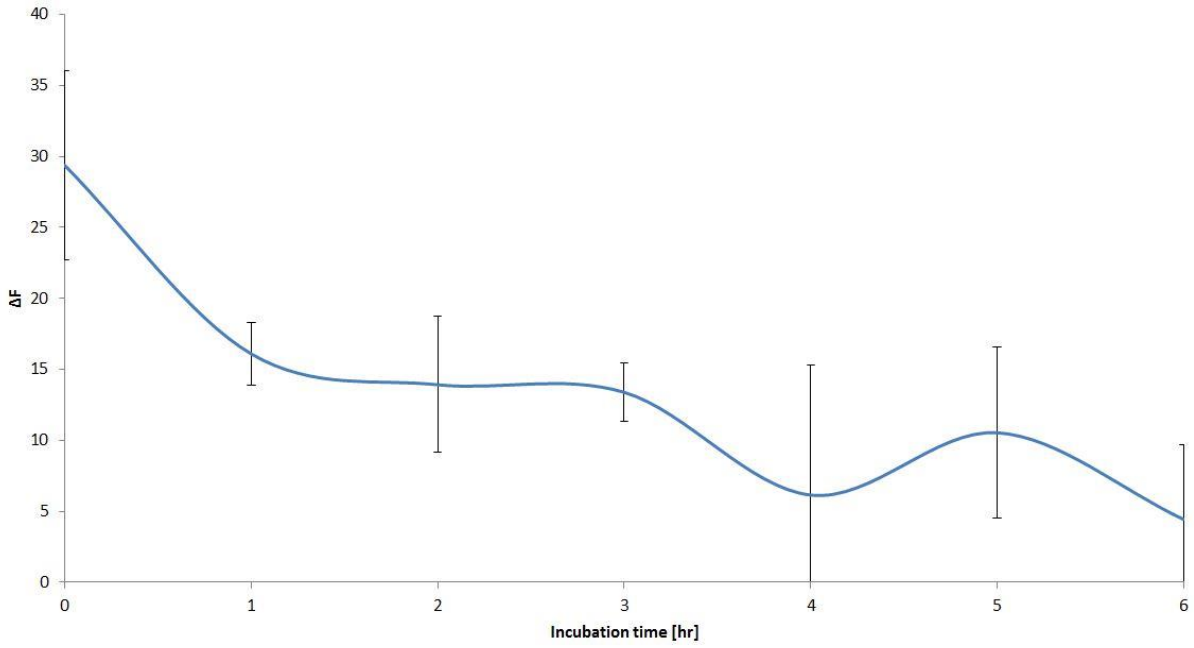


Figure 4.25 pSTAT3β_{TC} HTRF assay time course study (HTRF assay executed as described in 2.2.5. Means derived from N=3 +/- SD)

The normalized FRET signal (ΔF) decreased by approximately half after 1 hour equilibration and the signal continued to decrease over the six hour period. Components were added in the standard order (i.e. Protein, DNA fluorophores then KF) and the plate read immediately in order to limit signal degradation.

4.2.7 Determination of HTRF Assay Z Prime (Z')

Z prime (Z') is a statistical measure used for high-throughput screening assays to assess whether the difference between a positive and negative signal is large enough and reproducible enough. Z' requires four parameters in order to be calculated, these being the means (μ) and standard deviations (δ) of both the positive (p) and negative (n) controls (μ_p, δ_p, μ_n and δ_n). Z' is therefore defined as:

$$Z' = 1 - \frac{3(\delta p + \delta n)}{(\mu p - \mu n)}$$

A sample of 20 positive {biotinylated DNA and 20 negative (non-biotinylated DNA)} wells was set up, and the resulting Z' calculation is shown

Parameter	Value
μp	558.8917
δp	22.04244
μn	361.2222
δn	9.420832

Table 4.13 Means (μ) and standard deviations (δ) of both the positive (p = biotinylated DNA) and negative (n = non-biotinylated DNA) controls (μ_p, δ_p, μ_n and δ_n) for Z' data

$$0.52 = 1 - \frac{3(22.04 + 9.42)}{(558.89 - 361.22}$$

A Z' of less than 0 is deemed unacceptable for assay screening, a Z' between 0 and 0.5 is seen as acceptable and a Z' between 0.5 and 1 is seen as being excellent for high throughput screening. The Z' of 0.52 was good and the assay was further validated with the introduction of known STAT3 inhibitors.

4.2.8 Introduction of STA-21 STAT3 Inhibition

STA-21 was used as a known STAT3 inhibitor, but its mechanism of action is not fully understood, although it has been shown to inhibit STAT3 dimerisation formation and STAT3

DNA binding in separate studies (146). STA-21 was, therefore, a suitable positive control for inhibition of the HTRF assay signal, as shown in figure 4.32.

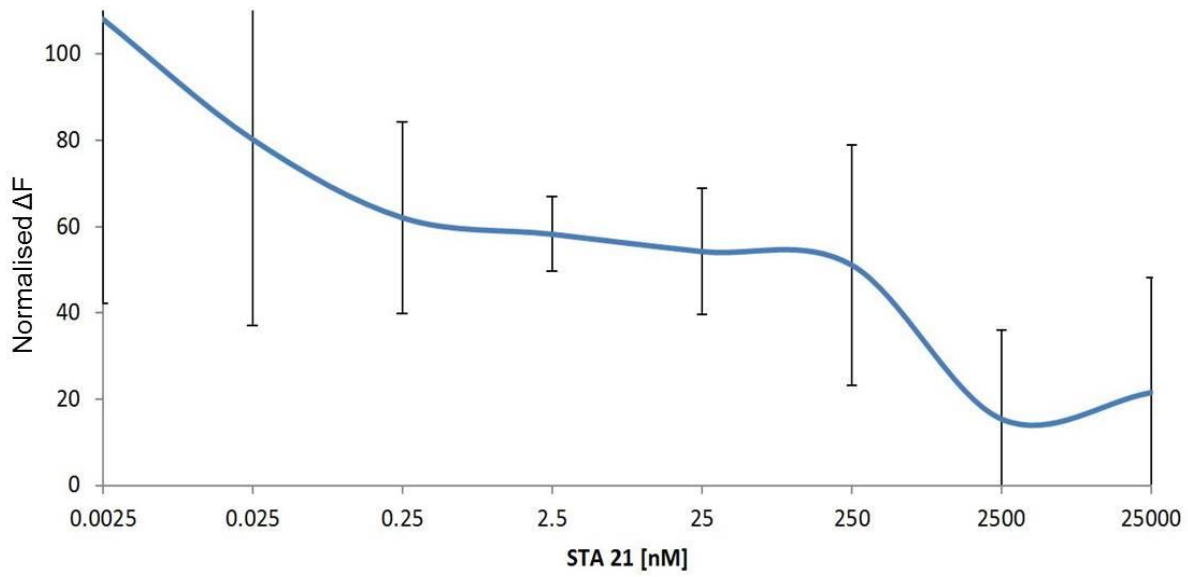


Figure 4.26 STA-21 inhibition titration P-STAT3β_{TC} HTRF assay. (HTRF assay executed as described in 2.2.5. Means derived from N=3 +/- SD)

The inhibition was normalized against the standard DMSO only control (cps665 nm/615nm)-mean cps DMSO)/cps DMSO) x100). A clear inhibition titration of signal is seen as the STA-21 concentration rises to its maximum of 25 µM. This acts as a proof of principle for the mode of action of the assay and also for its ability to be used to screen for STAT3-DNA binding inhibitors. This also suggests that STA-21's mechanism of inhibition of the STAT3 pathway in cancer cells is due to a local mechanism of action, through direct binding to either the STAT3 monomer or preventing dimer formation or through binding directly to the DNA sequence and preventing STAT3 binding.

4.2.9 Inhibition Study Utilising SH-2 Domain Mimic Peptides

Short sequence peptides that mimic the phosphorylated SH-2 domain of STAT3 (in and around phosphorylated tyrosine residue 705) were used as positive control inhibitors. These peptides have been regularly used to study STAT3 homodimerisation inhibition as they mimic the specific protein-protein interaction SH-2 domain of STAT3 (97, 135, 175). When present at high concentration the peptides bind to the SH-2 domains of STAT3 monomers rendering those proteins inert to dimer formation and subsequent DNA binding. Figure 4.33 displays the inhibition seen with 40 nM peptides in the 2:1 HTRF assay format (100 nM pSTAT, 50 nM DNA)

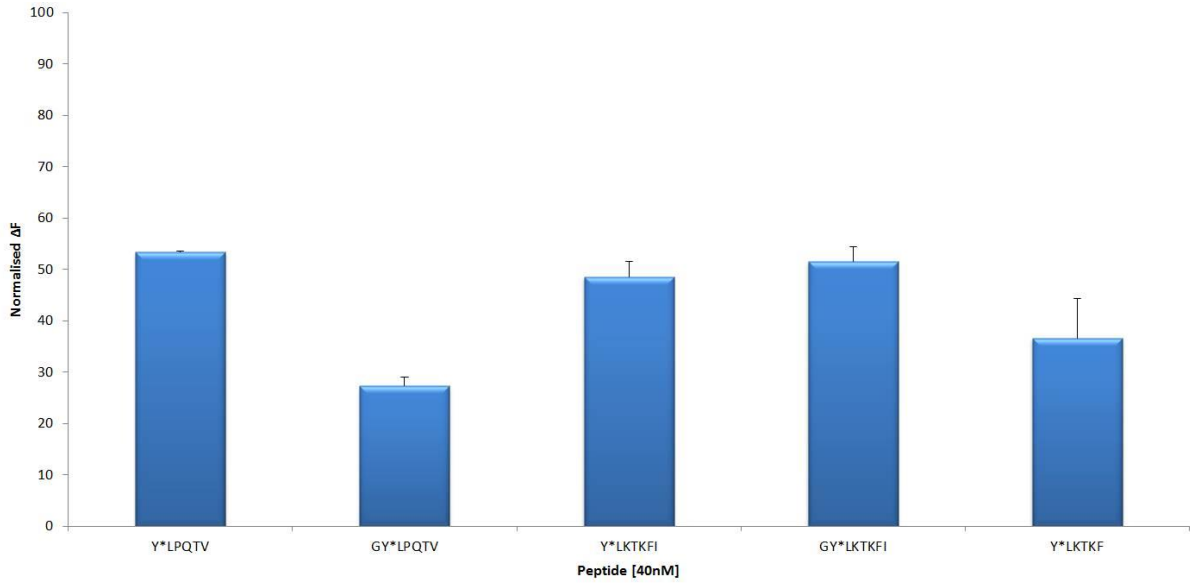


Figure 4.27 Inhibition using SH-2 domain mimic peptides at 40 nM. HTRF assay executed as described in 2.2.5. Means derived from N=3 +/- SD

All these phosphorylated peptides have been shown to inhibit STAT3 homodimerisation and this activity has been confirmed in this HTRF assay. LPQTV peptides were first shown to have an IC_{50} of 8 µM in a STAT3 mouse model (133). The level of inhibition (~50%) is in line with the ratio of peptide to protein (40 nM/100 nM) confirming that the mechanism of inhibition is as hypothesized.

5 Biological Screen of Compound Library

5.1 Introduction

As discussed in the chapter 1 Introduction, the STAT3 signalling pathway plays an important role in cell growth regulation, proliferation and apoptosis. The deregulation of this pathway in many human cancers however leads to increased malignancy and survival. STAT3 is therefore a greatly attractive target for cancer therapy.

The compound library described in this work resulted from the previously published RH06 (fig.5.36) (159). This was shown to have activity in a fluorescence polarisation assay, as well as specific cytotoxicity to MDA-MB-231 breast carcinoma cells when compared with A4 (STAT null) cells.

STAT3's protective effect in cancer cells is due to the overexpression of STAT3 target genes as the constitutively active STAT3 homodimerise and binds to promoter sequences of DNA. This in turn leads to expression of mRNA for survival proteins. It is this complex of STAT3 homodimer bound to DNA consensus sequence that we are looking to disrupt with a drug-like small molecule.

High throughput screening assays are an efficient way to narrow down libraries of inhibitors and spot links between important scaffolds and chemical groups. Assays of this type traditionally used a radioactive isotope as its signal, but recently fluorescence and luminescence signals have been preferred due to their safety and ease of use. The HTRF assay developed utilises the FRET phenomenon and reduces the quenching of signal effect. This assay is novel in the way it utilises a fluorophore on the DNA consensus sequence and fluorophores on the STAT3 homodimer. Therefore, the assay identifies inhibitors of the complete transcription complex, be they inhibitors of either STAT3 dimerization or STAT3-DNA interaction.

The basic structure scaffold for the compound library is seen in figure 5.36

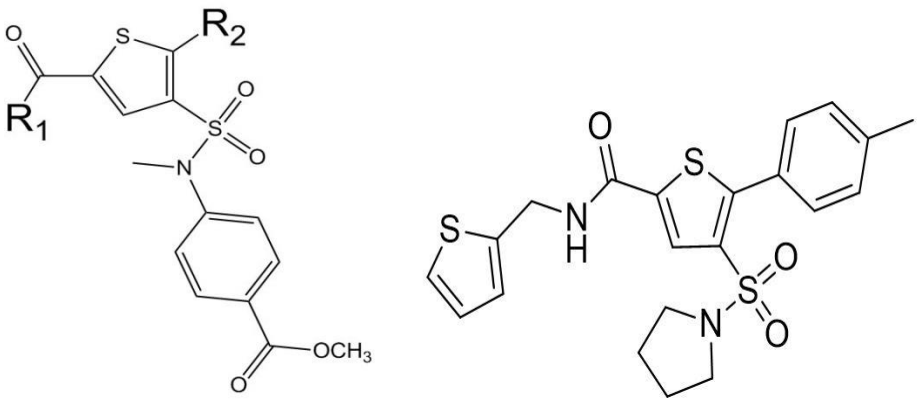


Figure 5.28 Basic scaffold of compound library alongside original RH-06 structure.

In the first generation of compounds R_2 represents a bromine residue and the R_1 groups of the 1st generation intermediates are shown in table 5.16.

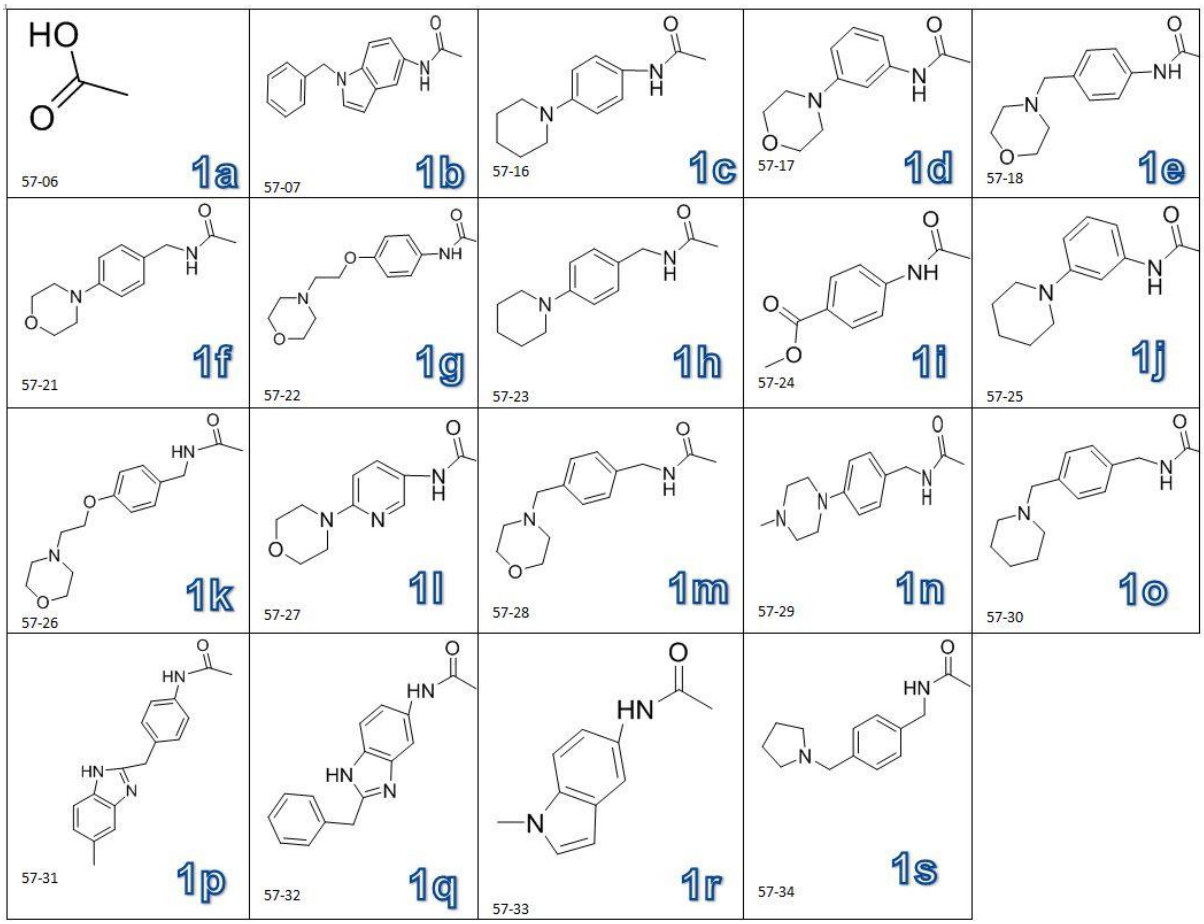


Table 5.14 19 1st generation intermediates with working and published nomenclature. (Only R^1 side chains shown, R^2 group was bromine.)

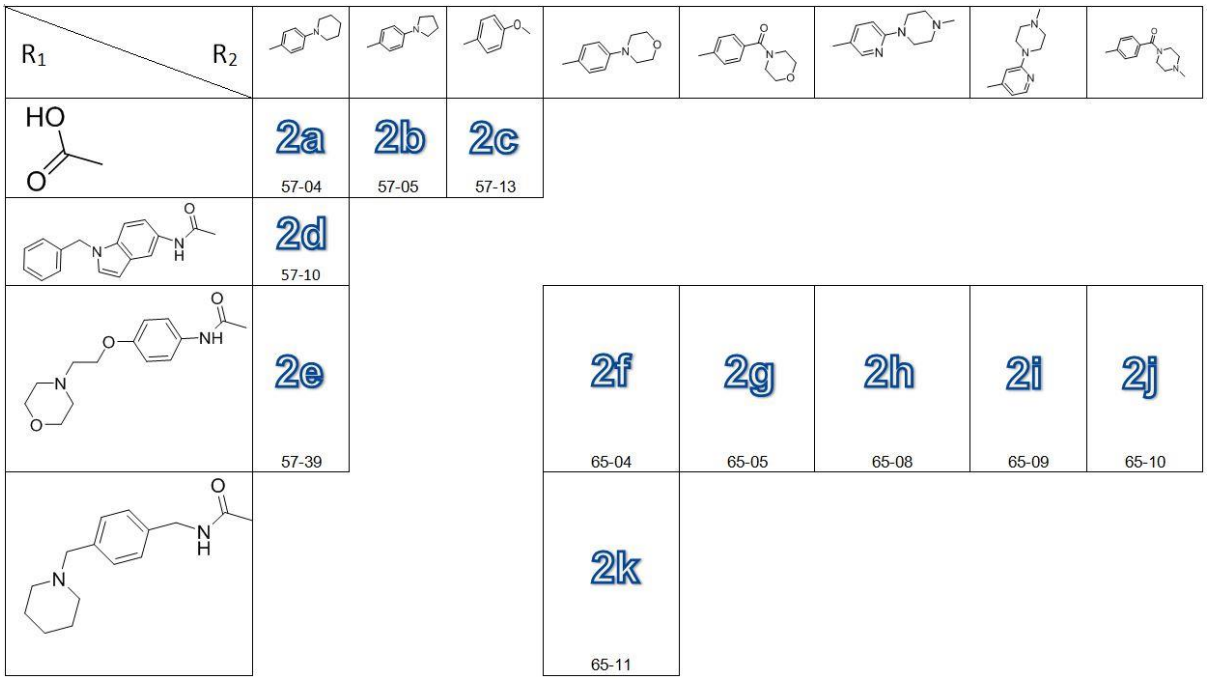


Table 5.15 2nd generation final compounds, R_1 groups shown on Y axis, R_2 groups shown on X axis, compounds published names in bold and laboratory names in small type.

The second generation of inhibitors take promising first generation intermediate R_1 and add novel R_2 in order to improve binding affinity to the SH-2 association region. Modelling simulations suggest that the R_2 moiety is essential for stable binding to residue P704 of the SH-2 domain.

5.2 Results and Discussion

5.2.1 Activity Study of Four 1st Generation Intermediates

Four 1st generation intermediates from the HTRF study were selected for cytotoxicity assay screen and endpoint PCR in order to link activity in the HTRF assay and cytotoxic effects to the compounds effects to downstream regulated gene profiles through PCR; this in turn would be linked to *in silico* modelling studies for the compounds.

The structures of compounds **1p, 1q, 1r** and **1s** are given in figure 5.37. The basic structure scaffold is followed by the side chain differences at the R'1 position.

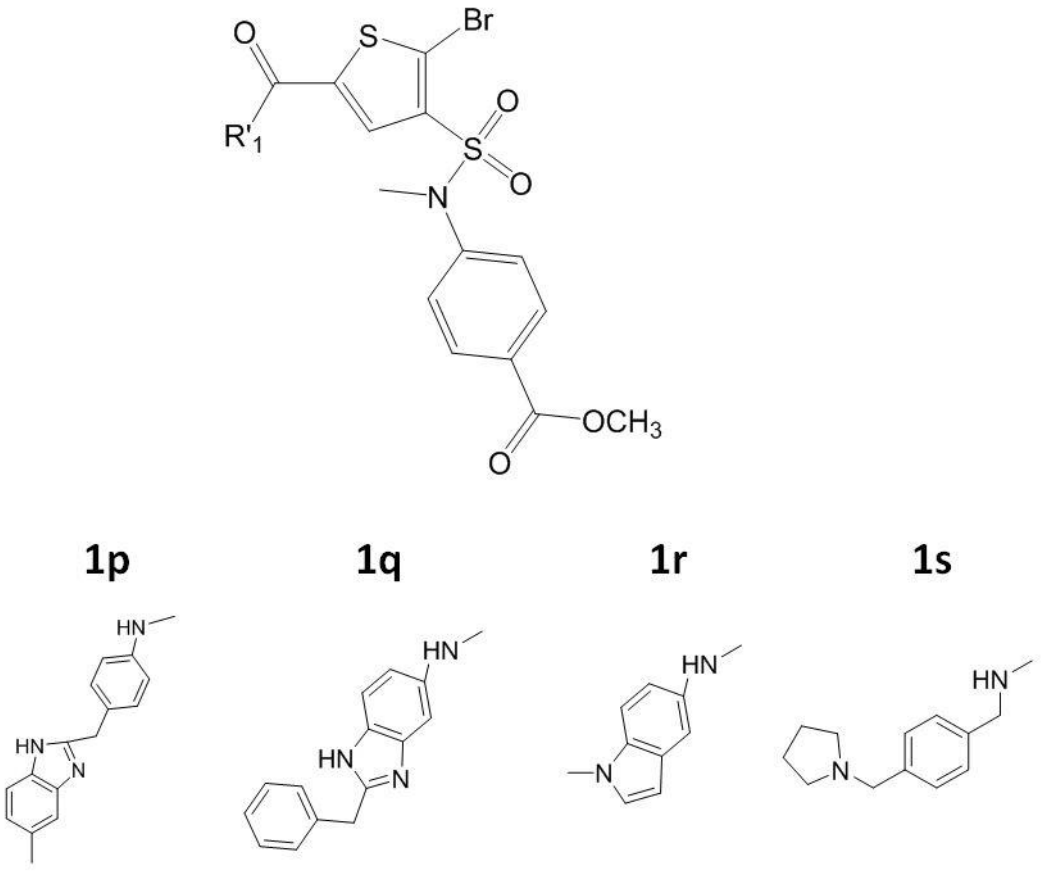


Figure 5.29 Basic intermediate structure with R'1 side chain differences

The compounds activity in the HTRF assay is summarised in figure 5.36. All compounds showed good inhibition with ≥ 50% inhibition at 40 µM except **1q**, which showed no significant inhibition compared with the DMSO control.

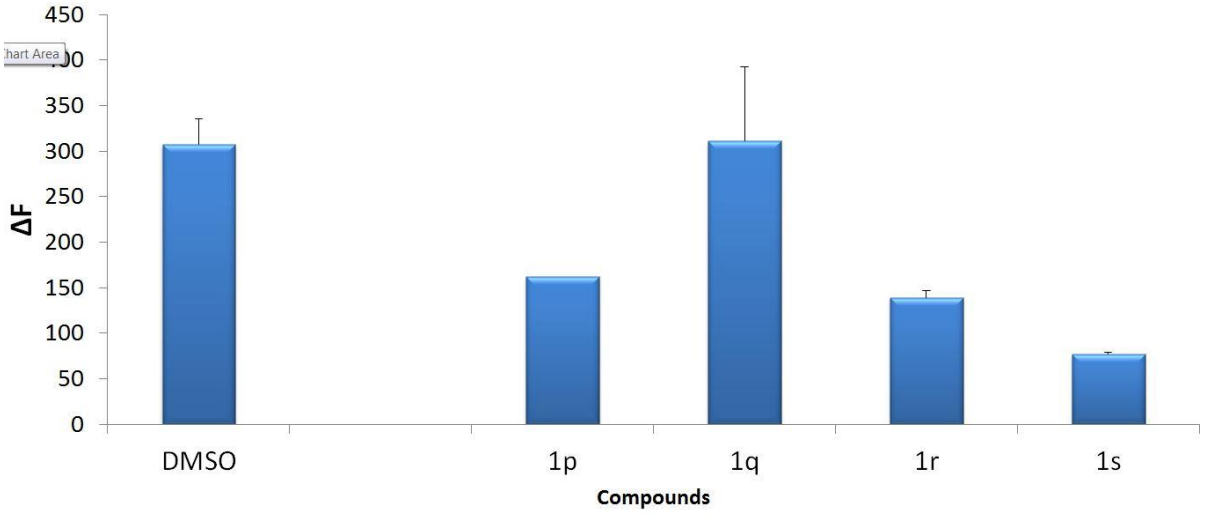


Figure 5.30 pSTAT3β$_{TC}$ HTRF (2.2.5.3) data summary for 1p, 1q, 1r and 1s (Means derived from N=3 +/- SD)

1p showed an inhibition of 48%, **1r** of 55% and **1s** of 74% compared with the DMSO control. **1s** showed a consistent inhibition with low variation in standard deviation and was considered the molecule of most interest for going into the cytotoxicity study.

The cytotoxic activity of the four compounds in MDA-MB-231 and A4 was compared at a concentration range between 1 and 100 µM.

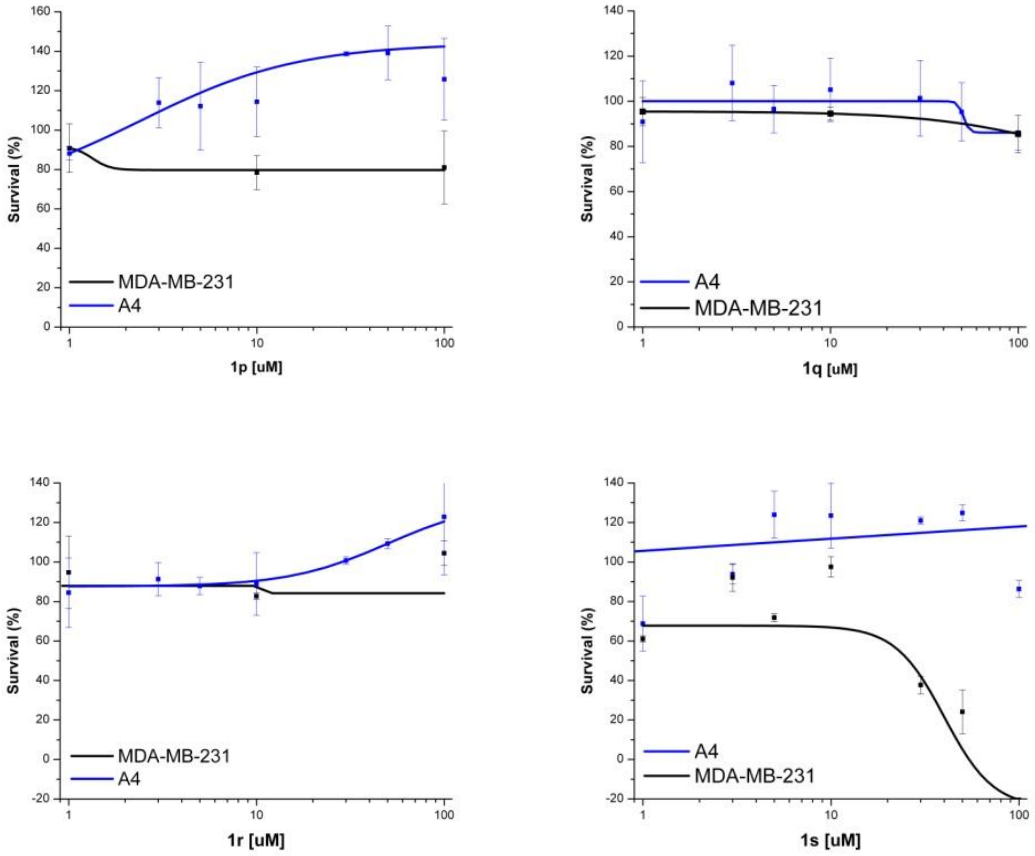


Figure 5.31 MDA-MB-231 and A4 MTT comparison (as described in section 2.2.6.5) (1p, 1q, 1r and 1s). Means derived from N=3 +/- SD.

Only **1s** showed any cytotoxicity with an IC_{50} of 25.9 µM in MDA-MB-231 breast carcinoma cells. This cytotoxic effect was STAT3 specific. However, no cytotoxic effect was seen in the STAT3 null A4 cell line (166). **1s'** activity from the HTRF assay was supported by a specific cytotoxic effect in MDA-MB-231 cells prior to studying the molecules activity on the downstream target genes expression of STAT3. Cytotoxicity data did not always correlate with the hypothesised activity from HTRF and endpoint PCR results, as can be seen with **1p** and **1r.** This was due to the compounds precipitating out in the MTT well conditions, often

seen as coloured crystals at the bottom of the test plate wells. Therefore the expected levels of drug concentrations could not have been achieved. However, the data are still presented here for comparison and completeness.

The downstream effects of the four compounds on the transcriptosome were assessed by PCR; the cell lysate of exposed MDA-MB-231 breast carcinoma cells was extracted and the mRNA isolated. The mRNA was then transcribed to cDNA by reverse transcriptase. Select genes were then multiplied using specific primers and polymerase chain reaction. The Endpoint PCR study was carried out in conjunction with Ms Julia Mantaj as per the protocol in section 2.2.7.

Figure 5.39 represents an endpoint PCR study using primers for STAT3, Cyclin D1, Bcl-2, and nicotinamide *N*-methyltransferase (NNMT), all genes that have been shown to be overexpressed in MDA-MB-231 cells (176). Glyceraldehyde 3-phosphate dehydrogenase (GAPDH) was used as a housekeeping control gene to check that downregulation of genes was STAT3 specific and not due to less specific downregulation. MDA-MB-231 cell lysate with and without lipopolysaccharide (LPS) were given as controls along with **1p, 1q, 1r** and **1s** as the test compounds.

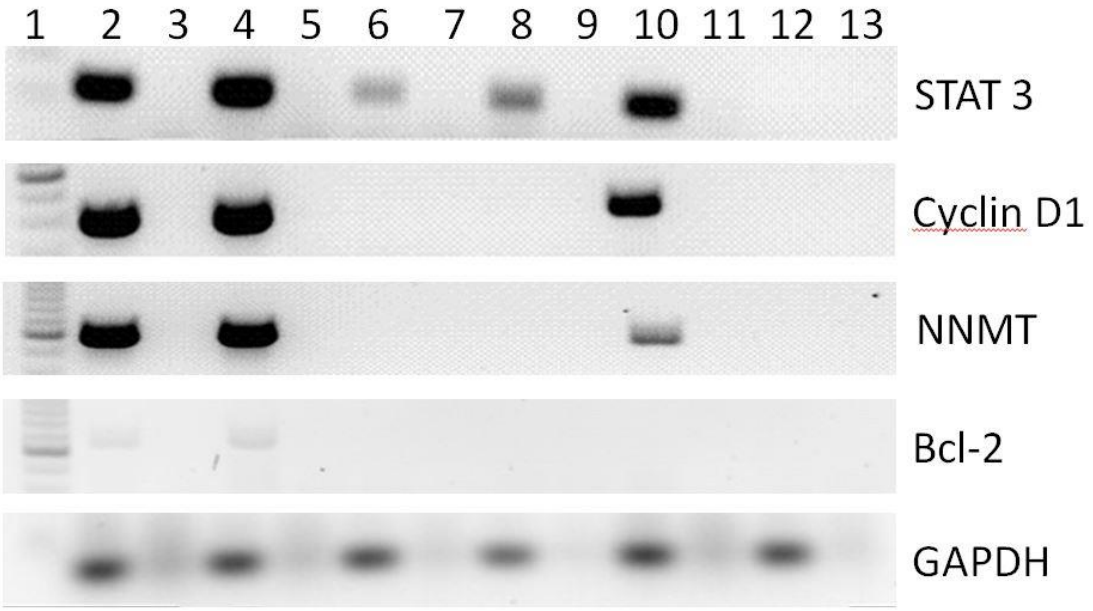


Figure 5.32 Endpoint PCR using primers for STAT3, Cyclin D1, NNMT, Bcl-2 and GAPDH. Lane 1 = 1Kb DNA ladder (NEB), Lane 2 = MDA-MB-231 cell lysate, Lane 3 = Lane 2 negative control, Lane 4 = MDA-MB-231 cell lysate and lipopolysaccharide (LPS), Lane 5 = Lane 4 negative control, Lane 6 = MDA-MB-231 cell lysate with LPS and 1p (100 µM), Lane 7 = Lane 6 negative control, Lane 8 = MDA-MB-231 cell lysate with LPS and 1q (100 µM), Lane 9 = Lane 8 negative control, Lane 10 = MDA-MB-231 cell lysate with LPS and 1r (100 µM), Lane 11 = Lane 10 negative control, Lane 12 = MDA-MB-231 cell lysate with LPS and 1s (100 µM), Lane 13 = Lane 12 negative control.

Lipopolysaccharide was active in its role of upscaling the STAT3 signaling pathway, as can be seen in the STAT3 PCR agarose gel (lanes 2 and 4), as it is known that active STAT3 upregulates its own transcription. **1s** (lane 12) was the most effective of the inhibitory compounds, which is consistent with the HTRF and cytotoxicity study. It knocked out the mRNA presence of all the genes except the GAPDH "housekeeping" gene. In fact, all compounds were specific enough to not affect the GAPDH level. Bcl-2 was knocked out by all compounds, but the background level was very low in LPS only. **1p** and **1q** (lanes 6 and 8 respectively) caused a total reduction in Cyclin D1, NNMT and Bcl-2 levels and greatly reduced the mRNA STAT3 level. **1r** (lane 10) was the least effective of the compounds, but

still caused a reduction in NNMT, Cyclin D1 and STAT3 levels. The high compound concentration (100 μM) makes comparison between the activity of compounds difficult, but a general proof of principle of mechanism of action can be deduced as to the ability of these compounds to inhibit the transcription profile within cancer cells.

The negative control lanes in the gel refer to the prior sample without the presence of Taq polymerase and were included in order to show efficacy of the enzyme.

5.2.2 Molecule Modelling Study on 1p, 1q, 1r and 1s

A docking study was undertaken on an average conformation of the STAT3 protein derived through a 10 ns explicit solvent molecular dynamics simulation, as described in section 2.2.6. In this study, **1p, 1q, 1r** and **1s** were docked in the SH2 binding domain and the GRID and GBSA binding scores were derived; each molecule was evaluated for its ability to prevent the interaction of Y705 and P704 of PYLKTKFI of homodimer B with the STAT3 monomer.

Molecule	GRID Score Protein (kcal/mol)	Interfere with P704 and Y705?	GBSA Score Protein (kcal/mol)	Interfere with Y705?
1p	-54.51	Y	-56.80	N
1q	-56.27	N	-61.69	Y
1r	-48.23	Y	-55.60	Y
1s	-54.31	N	-62.10	N

Table 5.33 GRID and GBSA score comparison for 1p, 1q, 1r and 1s

In biological experiments, **1s** was observed to produce cytotoxic activity in the STAT3-dependent cell line and was not cytotoxic in the STAT3-null cell-line (A4) indicating selective STAT3 inhibition. Although the molecule was not observed to interfere with binding of either Y705 or P704 in docking studies, it was seen to block binding of the central amino acids (LKT) of the PYLKTKFI amino acid chain to the octapeptide pocket. This interaction is characterised by the highly favourable GBSA score of -62.10 kcal/mol and GRID score of -

54.31 kcal/mol, and the binding is stabilised through the formation of a hydrogen bond between the sulphur of the thiophene group and nitrogen of Glu 639, and a second between the carbonyl of the sulfoxyl group of **1s** and a carbonyl of Thr 715. It is also evident that **1s** is tightly bound to the protein as the phenolic moiety of the molecule is buried deep within the protein architecture, surrounded on all sides by Ile 586, Leu 608, Phe 651, Pro 679 and Lys 680. Non-covalent interaction plays a prominent role in this interaction, indicated by the large (-56.05 kcal/mol) van der Waal's component of the GBSA score. It is likely that the affinity of **1s** for the protein, and thus its cytotoxicity, occurs as a result of the phenolic moiety of the molecule 'anchoring' the compound in the protein.

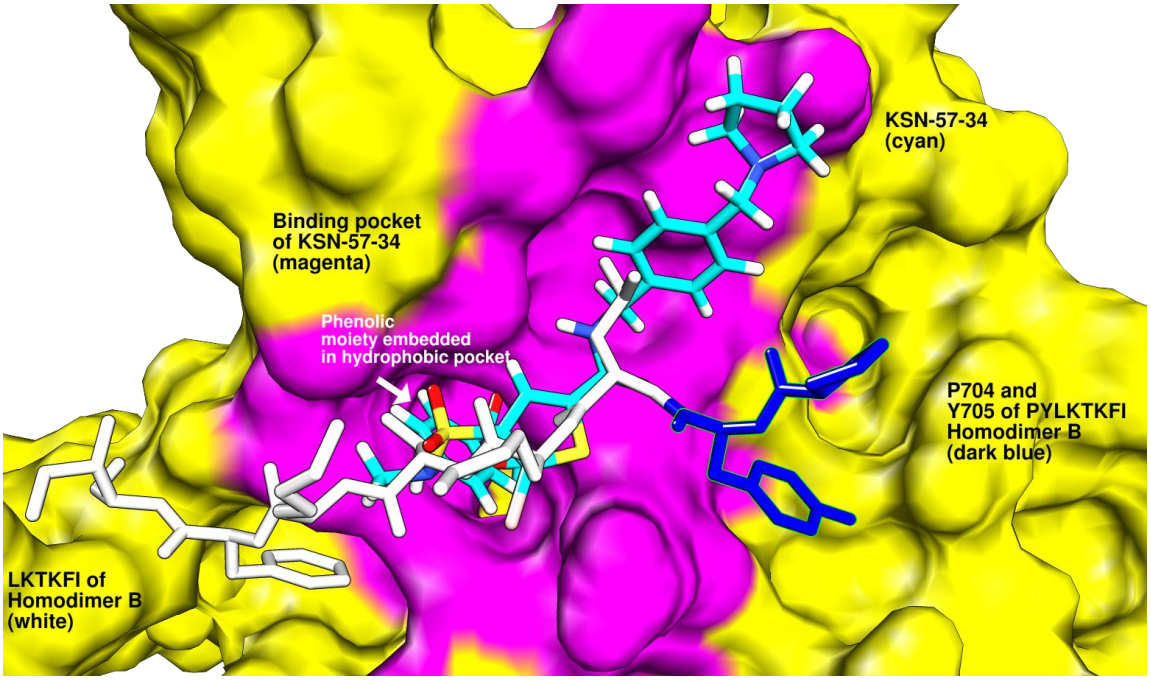


Figure 5.34 Modelling representation of 1s

Interestingly, **1q** produces a similarly high GBSA score, presumably affected by the formation of a hydrogen bond between the sulphur of Cys 713 and ring nitrogen of indole. In this respect, a high GRID score (-56.37 kcal/mol) also indicates snug accommodation of the compound in the SH2 domain. However, **1q** blocks PYLK from binding to its binding pocket, but does not embed in the protein in a similar manner to **1s**, and it is this difference that may

account for the lack of STAT3 binding disruption observed for **1q** relative to **1s**. Similarly, **1p** and **1r** also block interaction of PYLKTKFI with homodimer A (residues blocked are LKT and PYLKT, respectively), but do not embed in the protein in a similar manner to **1s**, reflecting the differences in protein cytotoxicity observed.

It is also interesting to note that the unsaturated fragments (for example indole and phenyl) of **1p, 1q** and **1r** are less favourable in activity than the unsaturated pyrrolodine. This may occur due to the kinked pyrrolodine fitting the protein cavity to a greater extent than the extended phenyl/benzofused building blocks.

5.2.3 Single Point Inhibitor Screen (100μM)

Employing the HTRF assay developed in Chapter 3, a screen of 19 compounds (15 1st generation and 4 2nd generation) was performed. These compounds were synthesized by our group and based on the structure of the pre-published inhibitor RH-06 (159). FRET signal was normalized against the background signal (the two fluorophore constructs) and ΔF calculated. Percentage inhibition compared with the DMSO control is shown in triplicate with the mean and standard deviation shown.

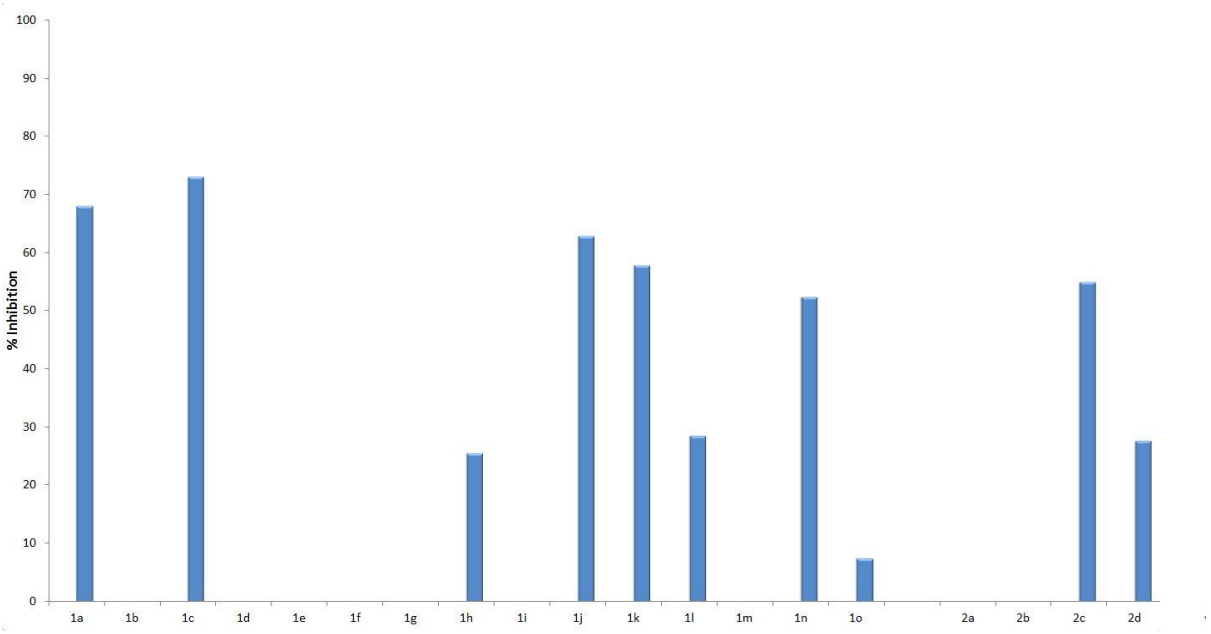


Figure 5.35 pSTAT3β$_{TC}$ HTRF (as described in 2.2.5.3) compound percentage inhibition at 100 μM (Means derived from N=3 +/- SD)

When compared with the DMSO only negative control, 10 of the 20 compounds tested showed an inhibition at 100 μM. Of the compounds that showed no inhibition compared with DMSO, this may be due to the compounds precipitating out of solution.

5.2.4 Front Runner Single Point Inhibitor Screen (10 µM)

The six compounds that showed greater than 50% compared with DMSO were screened again at a compound concentration of 10 µM. STATTIC is included as a positive control of known STAT3 inhibition.

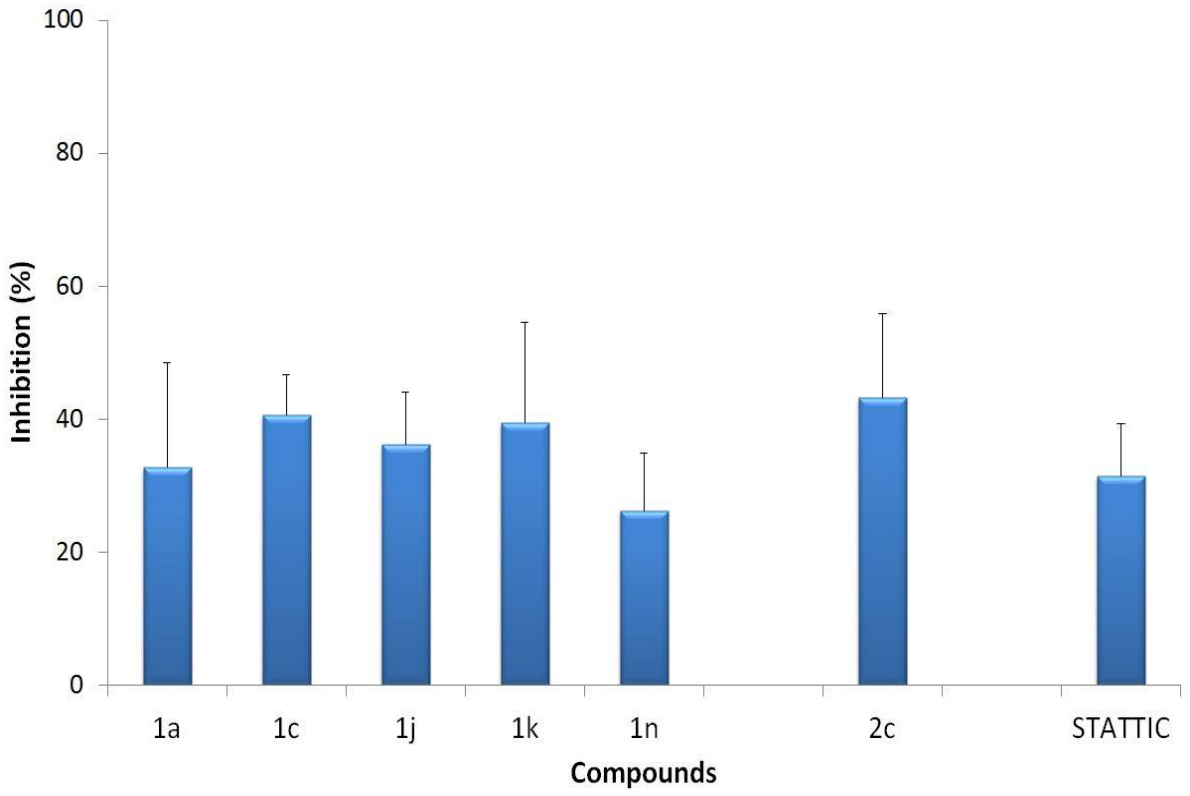


Figure 5.36 pSTAT3β$_{TC}$ HTRF (as described in 2.2.5.3) percentage inhibition comparison at 10 µM (Means derived from N=3 +/- SD)

There was a proportional inhibition shift, with all compounds showing a greater percentage inhibition at 100 µM than at 10 µM, as seen in table 5.20.

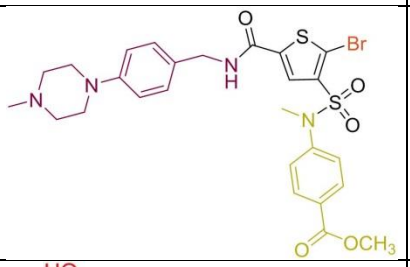
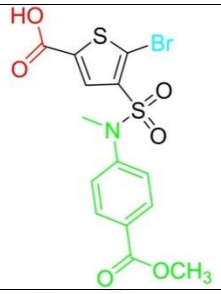
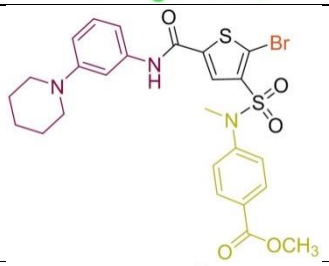
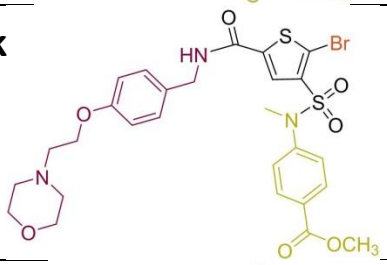
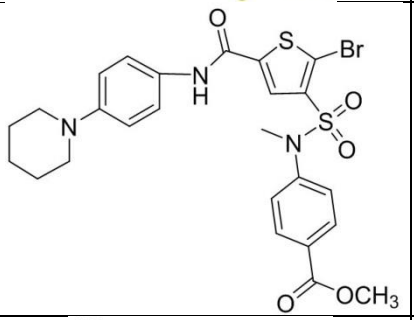
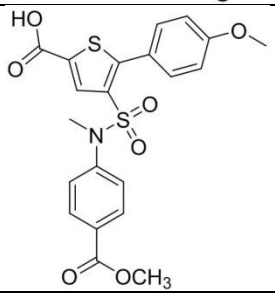
Compound structure	Inhibition (%) ranked at 10 µM	Inhibition (%) ranked at 100 µM
1n 	**27**	**52**
1a 	**33**	**68**
1j 	**37**	**63**
1k 	**39**	**58**
1c 	**40**	**73**
2c 	**43**	**55**

Table 5.16 Point inhibition (%) rankings at 10 µM and 100 µM

Table 5.20 shows the compound structures and ranks them according to their percentage inhibition compared with the DMSO only control at 10 and 100 µM point inhibitions. Five of the six molecules are intermediate 1st generation compounds with a bromine residue in the R'2 position. **2c**, however, is the first of the final compounds to show potential. These compounds have tolyl groups in the R'2 position and show greater binding affinity in molecular modelling studies.

The ranking of the compounds at 10 µM shows that **1n** is much more active at 10 µM than at 100 µM; this makes it a more attractive molecule for further study and the lead molecule from the initial screen.

5.2.5 Three Point Titration of the Initial Front Runners

A three point titration (25, 2.5 and 0.25 µM) was performed with the six lead molecules in order to investigate further the inhibitor concentration limits (Figure 5.44).

All compounds levelled off to their maximum inhibition levels, except **1k**, which showed a decrease up to 0.25 µM. A subsequent titration of **1k** showed that a maximum inhibition level of 5% was reached at 50 nM (data not shown).

These compounds were taken through to the cell culture assays to determine their availability and cytotoxicity within the cancer cell environment.

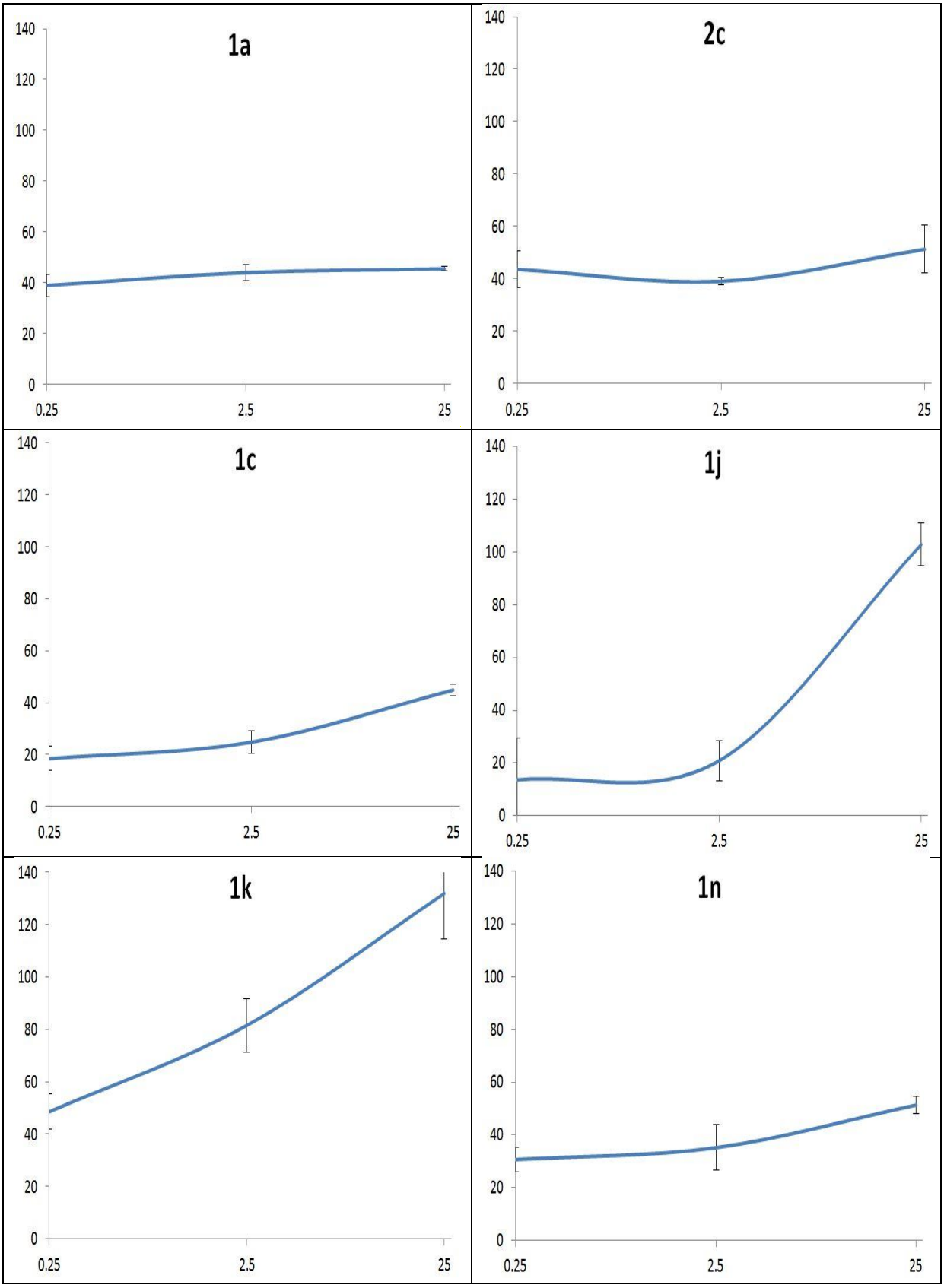


Figure 5.33 Three point titration of front runners (% inhibition vs compound concentration [µM])

5.2.6 Single Point Screen of Second Inhibitor Set

A second set of 10 compounds was tested in a 40 µM single point screen. **1p, 1q, 1r** and **1s** were additional 1st generation intermediates (with a bromine residue in the R'2 position).

Compounds **2e, 2f, 2g, 2h** and **2i** were derived from the **1g** intermediate and **2k** from the **1o** intermediate (see appendices).

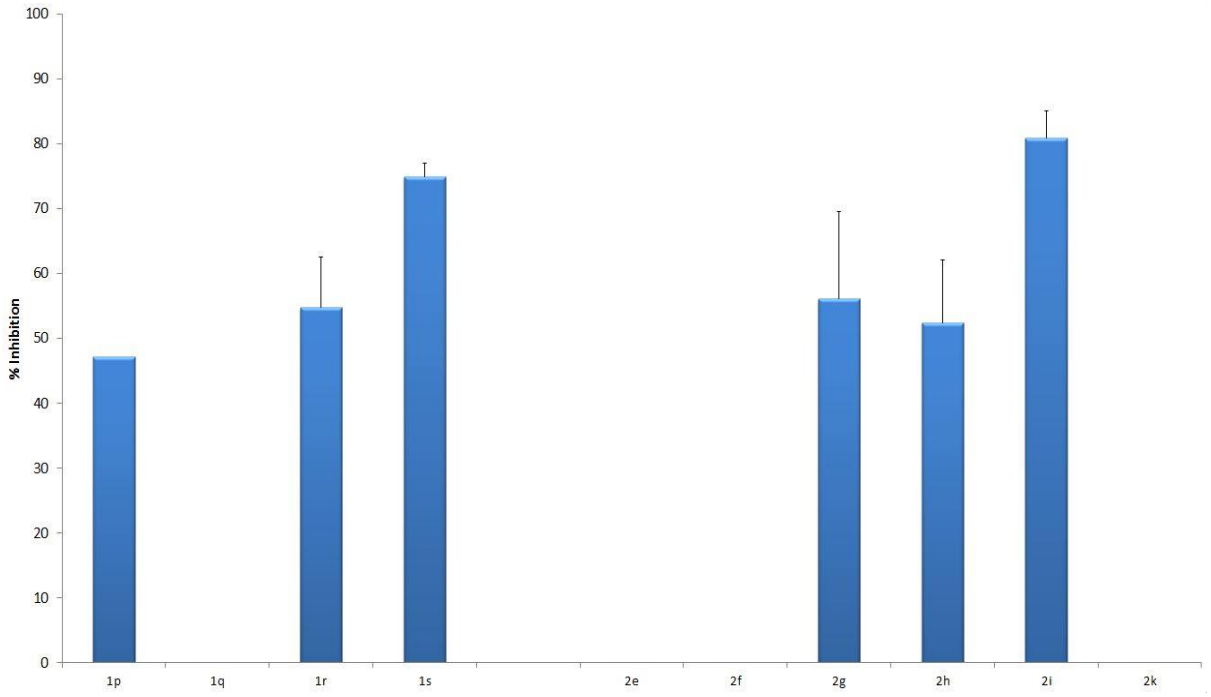


Figure 5.34 pSTAT3β$_{TC}$ HTRF (as described in 2.2.5.3) point inhibition results at 40 µM (Means derived from N=3 +/- SD)

Of the ten compounds, six showed a decrease in ΔF compared with the DMSO only control. These six compounds' inhibition profiles are summarised in Figure 5.45 and Table 5.21.

2i is the lead compound of the final compounds with an 81% inhibition of ΔF signal compared with the DMSO only control.

Figure 5.45 shows the six compounds that inhibited at 40 µM as a percentage inhibition compared with the DMSO only control.

Twelve of the original 29 compounds in the library showed potential in the HTRF screening assay. The next step (detailed in Chapter 6) was to screen the library for their cytotoxic effect in cancer cell lines. By correlating these results, along with the molecular modelling data, further a more accurate picture of the inhibitory nature of these compounds will come into focus.

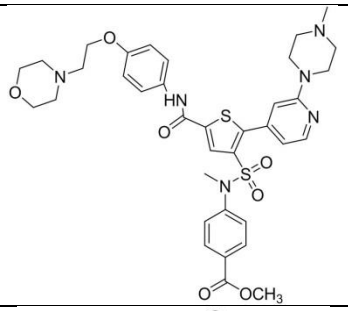
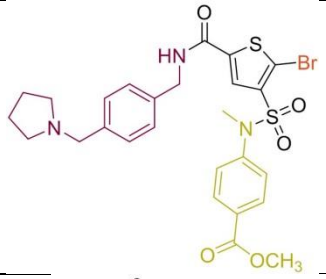
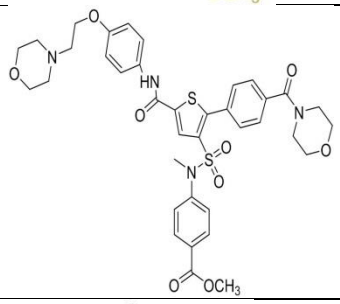
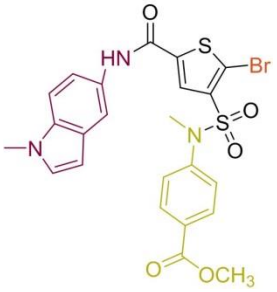
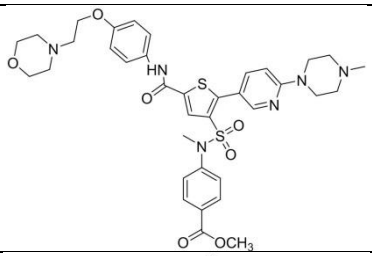
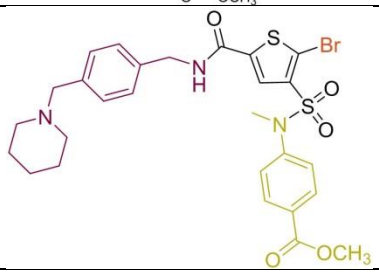
Compound structure	Inhibition (%) Ranked
2i 	**80.9**
1s 	**74.9**
2g 	**56.1**
1r 	**54.7**
2h 	**52.4**
1p 	**47.1**

Table 5.17 Point inhibition rankings (40 μM)

6 Effects on STAT3 Mediated Cellular Functions

6.1 Introduction

Chapter 5 detailed the screening of a targeted library of 29 compounds in a validated HTRF assay. In this chapter the same library of compounds plus some additionally developed final compounds were tested for their intercellular cytotoxicity using a MTT assay in four cancer cell lines. Results are shown for cell viability in HeLa, MDA-MB-231 and NCI-H1975 cancer cell lines. A STAT3 null cell line (A4) was used as a negative control to ensure that cytotoxicity was specific to STAT3 inhibition, as hypothesised.

Control inhibitors (Stattic and STA-21) are shown in MDA-MB-231 and A4 cells to validate assay specificity. The 1st generation intermediate front runners were then studied for cytotoxicity in HeLa and MDA-MB-231 cells. Second generation final compounds are then shown in MDA-MB-231 cells before moving on to a study of four targeted 1st generation compounds (**1p, 1q, 1r** and **1s**). These compounds are shown through HTRF pSTAT3β_{TC} inhibition, MTT cancer cell cytotoxicity and finally an endpoint PCR study of the effect of the compounds on downstream gene targets (STAT3, Cyclin D1, NNMT, Bcl-2) simulating therapeutic efficacy. Second generation front runners from HTRF pSTAT3β_{TC} inhibition assay are then progressed to MDA-MB-231 and A4 cells for cytotoxicity study, before a study of the three compounds of interest that were not active in HTRF, but had interesting cytotoxicity profiles (**2e, 2f** and **2k**).

Both generation front runners are then tested in NCI-H1975 non-small cell lung carcinoma cells, before a summary of all cytotoxicity profiles. *In silico* modelling results are then discussed and analysed. NCI-H1975 cells have constitutively active STAT3 activation due to a mutation in the epidermal growth factor receptor (EGFR) (177).

6.2 Results and Discussion

6.2.1 Analysis of MTT Data

As discussed in section 2.2.6.5., a transparent 96 well plate was seeded with cells in all wells except A1-6. The plate was then set-up as seen in figure 6.46 with triplicates of compound condition.

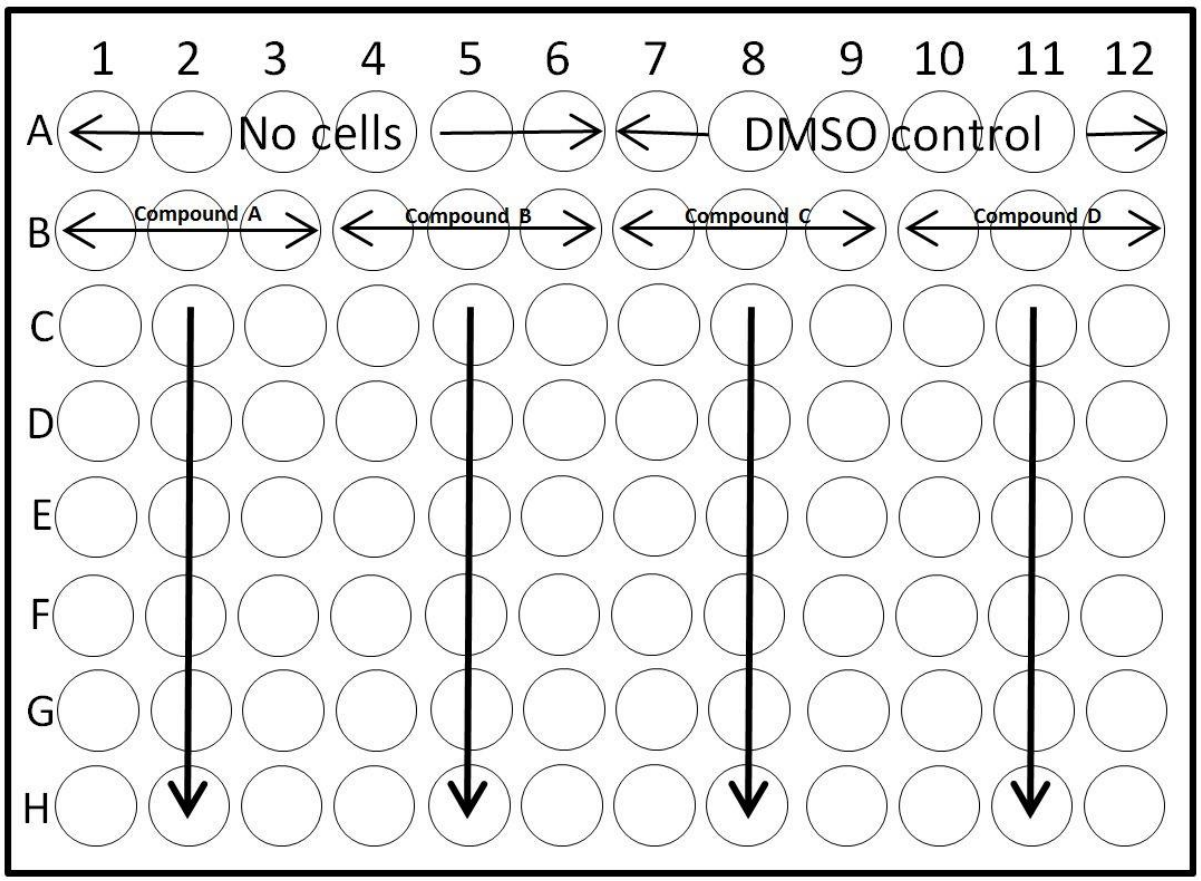


Figure 6.35 96 well template of MTT assay set-up. Row B compound concentration = 100 µM, C = 50 µM, D = 30 µM, E = 10 µM, F = 5 µM, G = 3 µM and H = 1 µM.

Once the experimental procedure was completed and the absorbance at 570 nm for each well obtained, an average of wells A1-6 and A7-12 was calculated (No cell background, $0_{Back,}$ and DMSO only control, 0_{DMSO}, respectively). Each well was then normalised using the following equation, where "*a"* is the pre-normalised absorbance at 570 nm.

$$X = \frac{(a - 0_{Background})}{(0_{DMSO} - 0_{Background})} \times 100$$

A mean and standard deviation of the normalised triplicates were then plotted on scatter charts using the origin to calculate IC_{50}s.

6.2.2 Positive Control Inhibitor STATTIC and STA-21

The cytotoxicity of known STAT3 inhibitor, STATTIC was tested in both the STAT3 dependent breast cancer cell line (MDA-MB-231) and the STAT3 null cell line (A4) as a positive control and an indication of the survival difference between the two cell lines within the concentration range 1-100 µM (figure 6.47).

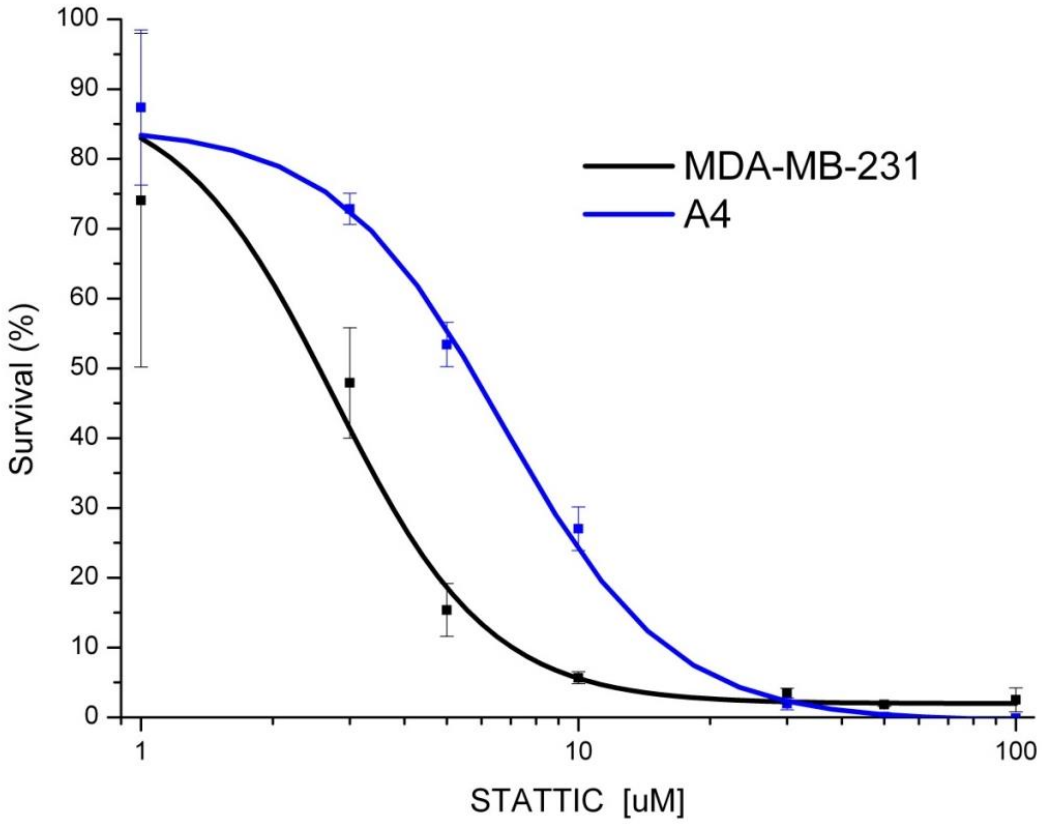


Figure 6.36 MDA-MB-231 vs A4, MTT cytotoxicity (as described in 2.2.6.5) in the presence of STATTIC (Means derived from N=3 +/- SD)

IC_{50} in MDA-MB-231 cells was ~2.5 µM, compared with ~5.7 µM in A4 cells. This decrease of over 55% in the STAT3 dependent cells represented a good bench mark to base subsequent decreases in survival with test compounds.

The apoptotic potential of the known STAT3 inhibitor, STA-21 (146) was tested for cytotoxicity using the STAT3 dependent and null cell lines.

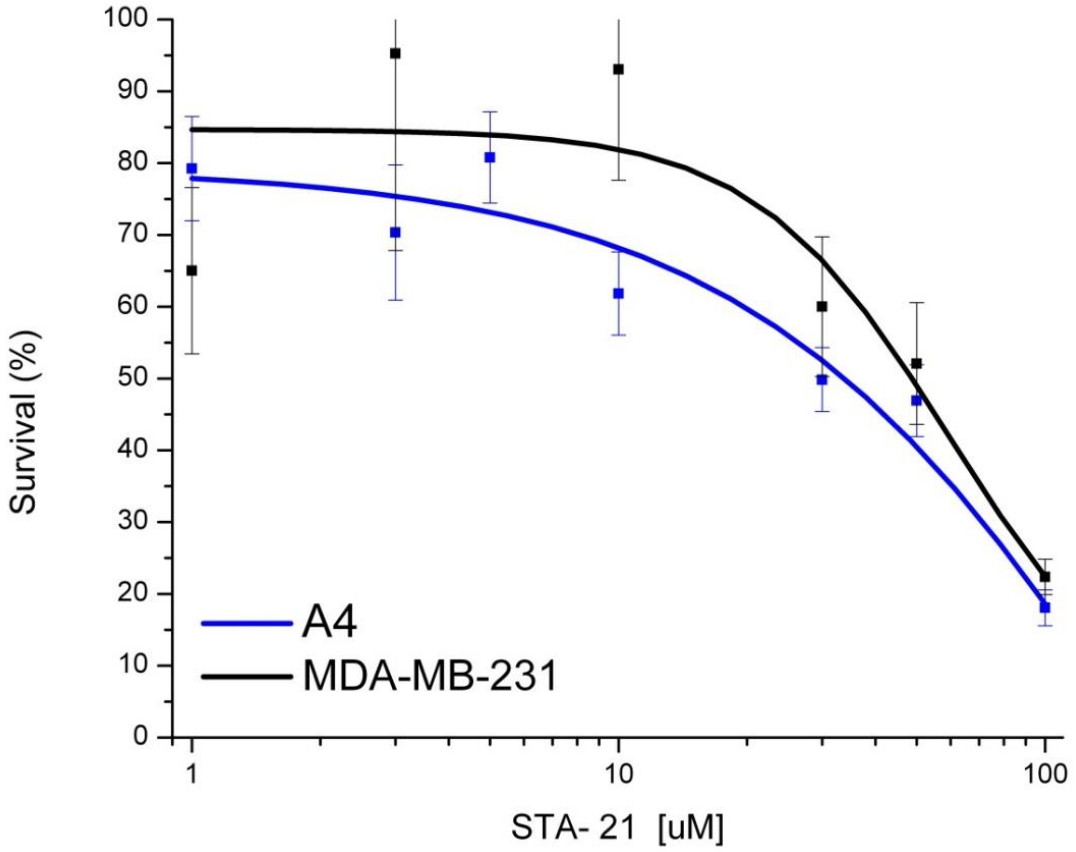


Figure 6.37 MDA-MB-231 vs A4, MTT cytotoxicity (as described in 2.2.6.5) in presence of STA-21 (Means derived from N=3 +/- SD)

The level of cell metabolic activity in MDA-MB-231 breast cancer cells was in line with the previously published data (Song, 2005). A 30% increase in cellular metabolic activity was found at 20 and 30 µM STA-21, and at those concentrations we see a comparable decrease in survival (20-30%).

Although active on STAT3, STA-21 clearly affects other key survival pathways. The IC_{50} in MDA-MB-231 cells was ~49 µM, compared with ~32 µM in A4 cells. STA-21 was more potent in the STAT3-null cell line displaying the additional mechanisms of action of STA-21, such as inhibition of TNF-α, IL-6 production and mRNA regulation of many factors (NF-κB, p65, JAK1, STAT3 and others)(147). Interestingly, this decrease was more than matched in the A4 cell line suggesting that the cytotoxic effect may not be STAT3 dependent.

6.2.3 Cytotoxicity Study on Lead 1st Generation Compounds

The five front running first generation compounds from the HTRF screen were tested for their cell cytotoxicity. An initial screen in STAT3 dependent cell line HeLa was performed with an inhibitor range of 1 µM to 100 µM.

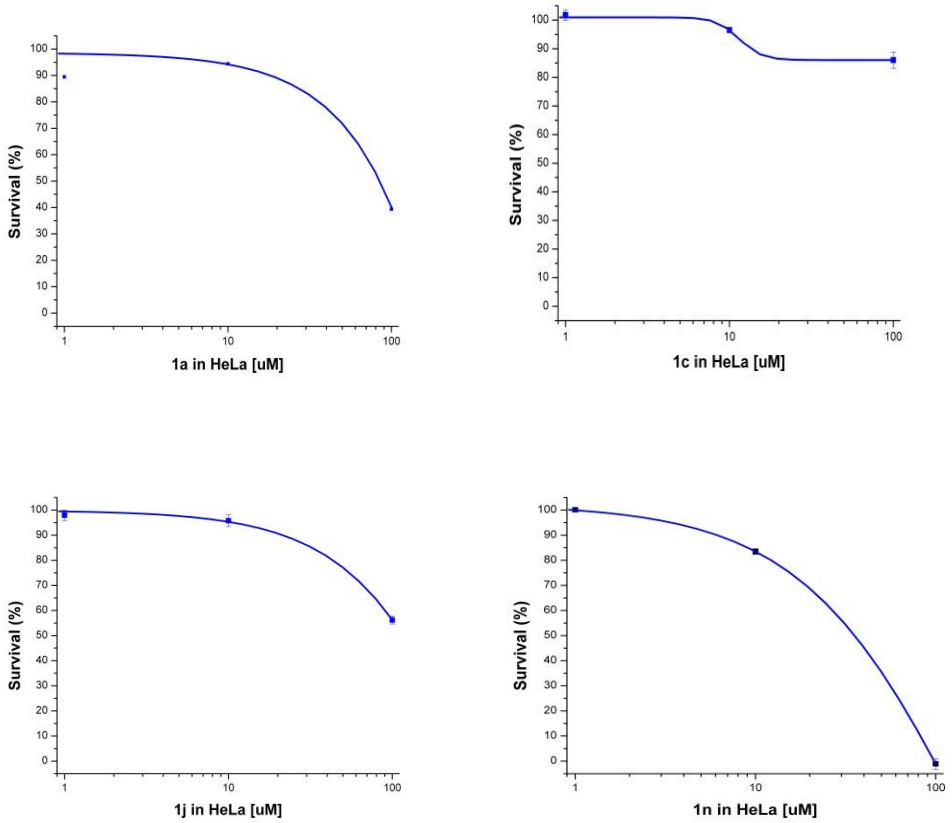


Figure 6.38 HeLa MTT cytotoxicity screen (as described in 2.2.6.5) of 1st generation front runners (Means derived from N=3 +/- SD).

1n was by far the most cytotoxic compound killing 100% of HeLa cells at 100 µM. A correlation between the HTRF assay and this screen is seen with **1n, 1a** and **1j** coming first, second and third, respectively, in both the HTRF screen and this MTT screen. **1k** data not shown as a data curve was not seen.

The 1st generation front runners were then tested for cytotoxicity in MDA-MB-231 breast carcinoma cells with results compared with HeLa cells.

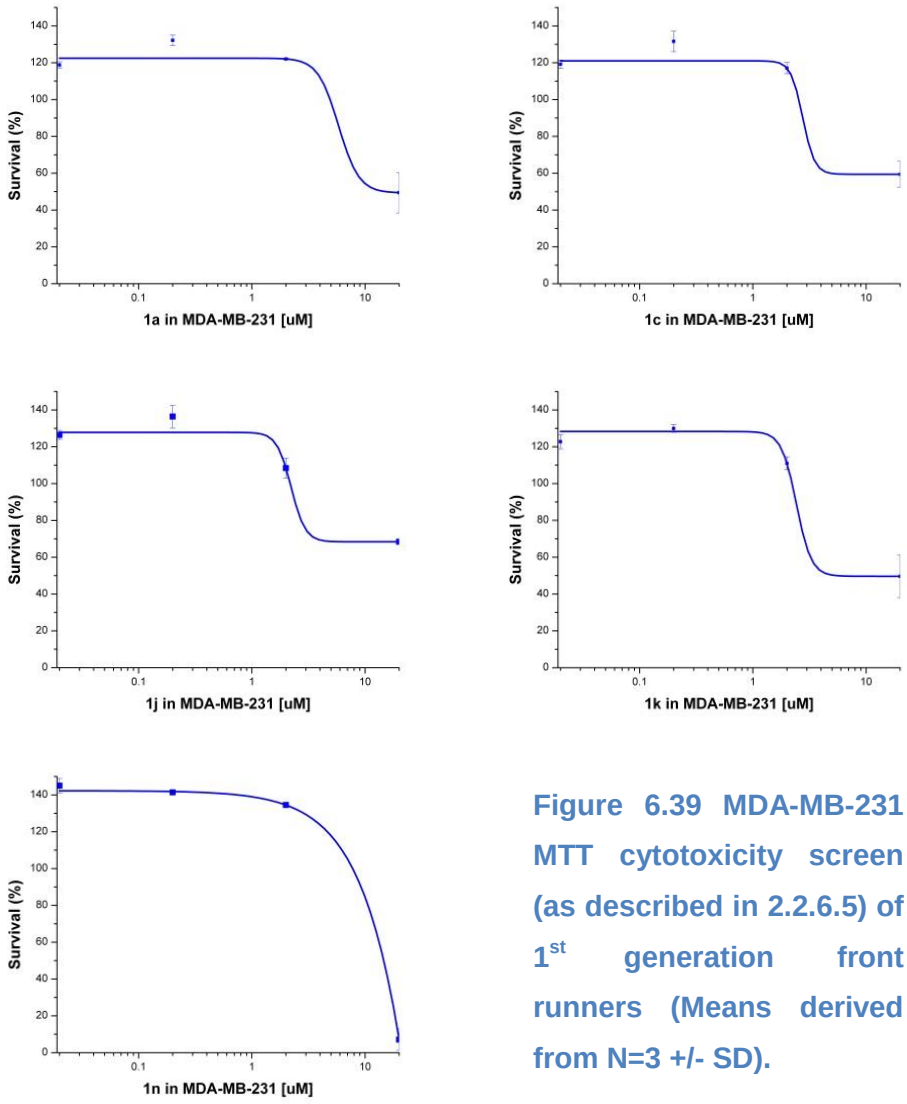


Figure 6.39 MDA-MB-231 MTT cytotoxicity screen (as described in 2.2.6.5) of 1st generation front runners (Means derived from N=3 +/- SD).

1n was the most cytotoxic compound in both cell lines and killed over 90% of MDA-MB-231 breast cancer cells at 20 µM. The compounds were more active at low concentrations in the breast cancer cells, which was encouraging due to MDA-MB-231 high STAT3 dependency.

1n was tested in MDA-MB-231 breast carcinoma cells, as well as in STAT null A4 cells, in order to test the specificity of the toxic effect and provide evidence for it being due to STAT3 inhibition.

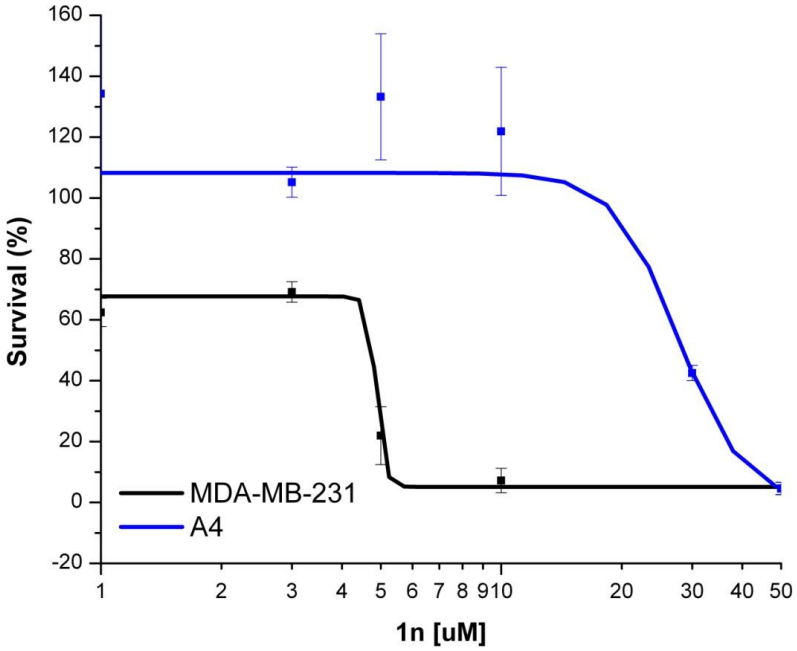


Figure 6.40 MDA-MB-231 vs A4 in MTT cytotoxicity screen (as described in 2.2.6.5) in the presence of 1n (Means derived from N=3 +/- SD)

An IC_{50} of 4.7 µM was seen in MDA-MB-231 compared with 28.2 µM for A4 cells, a specific decrease of 83%. When this result is considered alongside the HTRF data, **1n** was considered as the lead inhibitor of the first generation screen. An extrapolation of the MDA-MB-231 curve to the baseline of 100% survival would show a preferable sigmoidal curve. There may not have been the number of viable attached cells expected in the well, but the percentage cell death of those cells present is still significant.

6.2.4 Cytotoxicity Study with 2nd Generation Final Compounds

The four front runners from the HTRF 2nd generation screen were tested for their cytotoxicity in both MDA-MB-231 and STAT3 null A4 cells

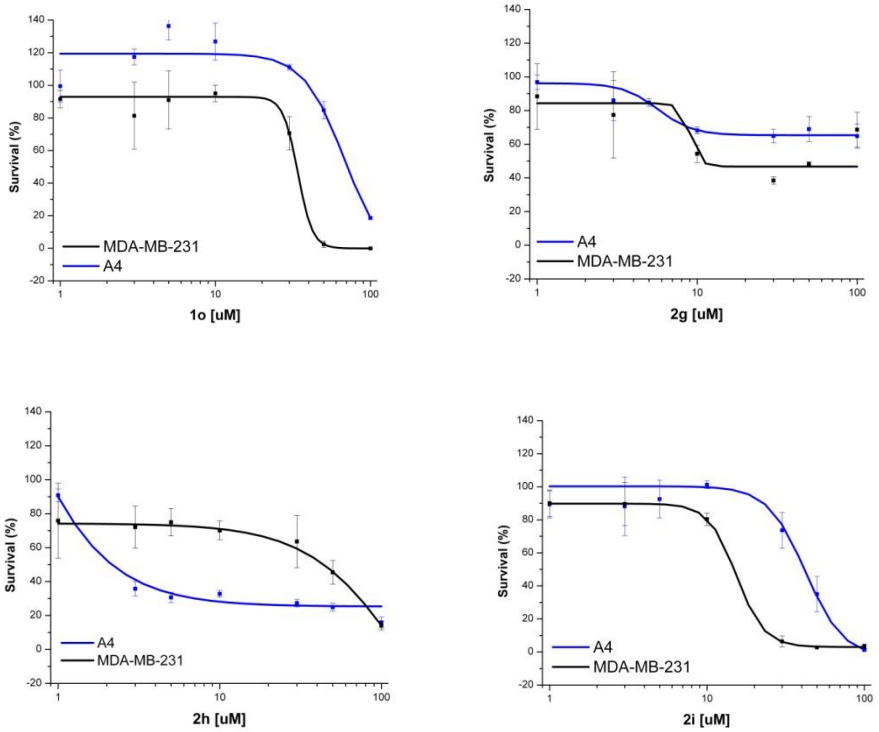


Figure 6.41 MDA-MB-231 vs A4 MTT cytotoxicity screen (as described in 2.2.6.5) in the presence of four 2nd generation front runners (Means derived from N=3 +/- SD)

2i was the most potent compound with an IC_{50} of 14.8 µM in MDA-MB-231 breast carcinoma and 41.4 µM in A4 cells showing a nice selectivity for the STAT3 dependent cell line over the STAT null one.

1o also showed potency and specificity (MDA-MB-231 = 32.7 µM, A4 = 70.1 µM) with an IC_{50} of less than 50% in MDA-MB-231 compared with A4.

2g showed very little increase in metabolic activity as the concentration was increased across both cell lines and a look at its effect in the modelling simulation is required to

determine whether the HTRF result is an anomaly. The best fit curve is included for clarity, but no IC_{50}s were calculated.

Similarly, **2h** showed an increase in metabolic activity with concentration, but the specificity is lacking. The compound was more active in A4 cells than MDA-MB-231 and, therefore, no IC_{50}s were calculated.

6.2.5 Cytotoxic Hit Compounds Un-Flagged by HTRF

Three compounds showed interesting preliminary cytotoxicity in the screens with MDA-MB-231 and A4 cell lines. These results are presented as they are to be carried forward into model screening studies.

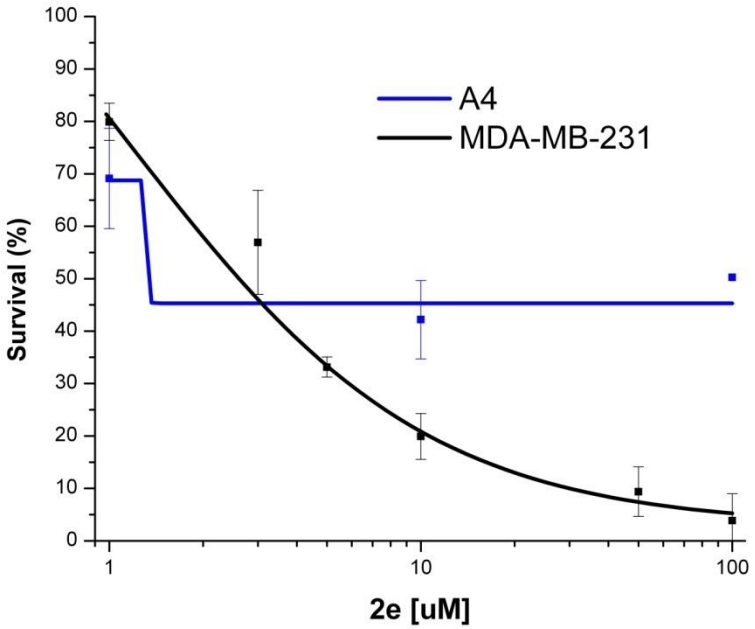


Figure 6.42 MDA-MB-231 vs A4 MTT cytotoxicity screen (as described in 2.2.6.5) in presence of 2e. (Means derived from N=3 +/- SD)

2e has a smooth titration with an IC_{50} in MDA-MB-231 cells of 2.6 µM. The IC_{50} achieved with this compound is the most potent seen in MDA-MB-231 cells and the consistency and reproducibility seen in the titration makes this compound a strong candidate for further investigation, despite a poor showing in the HTRF assay. This inhibition is also specific as there is no titration of metabolic activity with A4, 100 µM maintaining 50% cell survival.

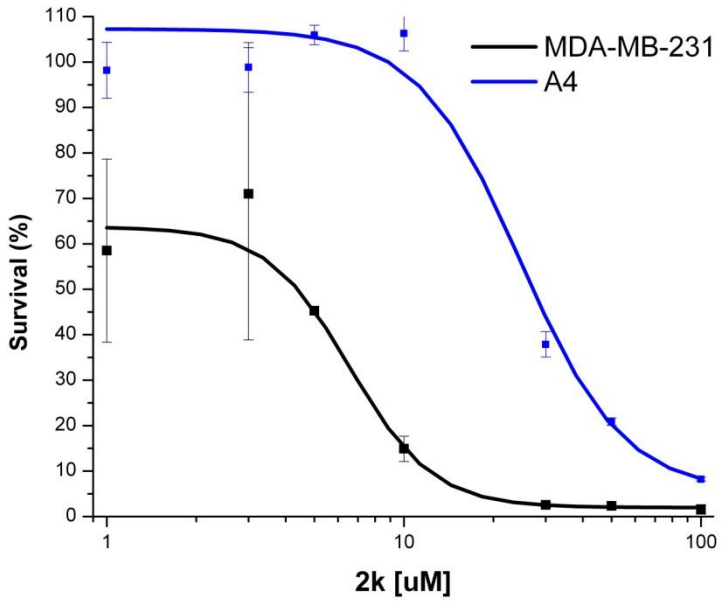


Figure 6.43 MDA-MB-231 vs A4 MTT cytotoxicity screen (as described in 2.2.6.5) in the presence of 2k. (Means derived from N=3 +/- SD)

2k had an IC_{50} of 4.3 µM in MDA-MB-231 cells compared with an IC_{50} of 27 µM in the STAT3 null A4 cell line. **2k** has the greatest differential between MDA-MB-231 and A4 IC_{50}, of all the compounds tested, with the A4 value being over six times greater than the MDA-MB-231 IC_{50}. This marks **2k** as a top hit compound despite its poor showing in the HTRF assay. Further in cell testing is required to elucidate the mechanism, although this compound was in the top ten *in silico* modelling ranking.

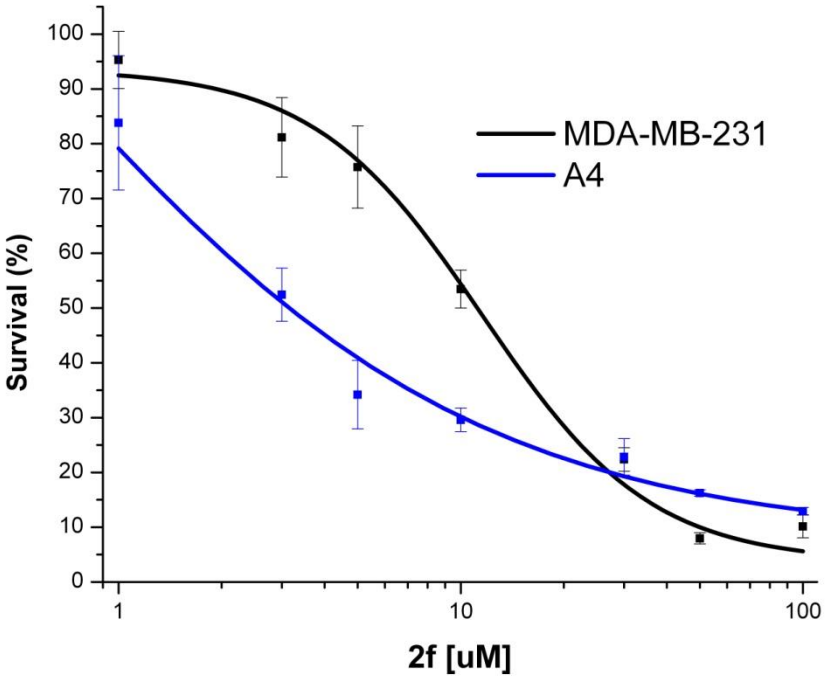
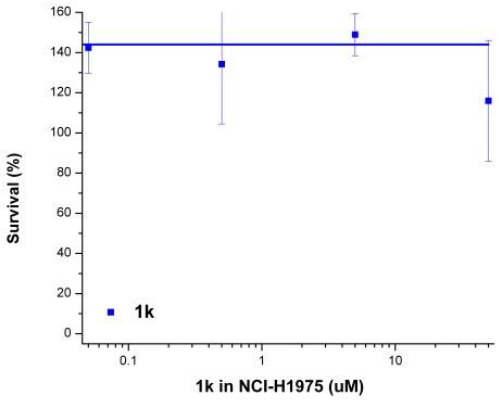


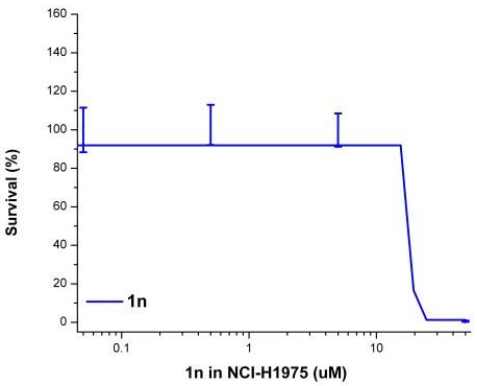
Figure 6.44 MDA-MB-231 vs A4 MTT cytotoxicity screen (as described in 2.2.6.5) in the presence of 2f. (Means derived from N=3 +/- SD)

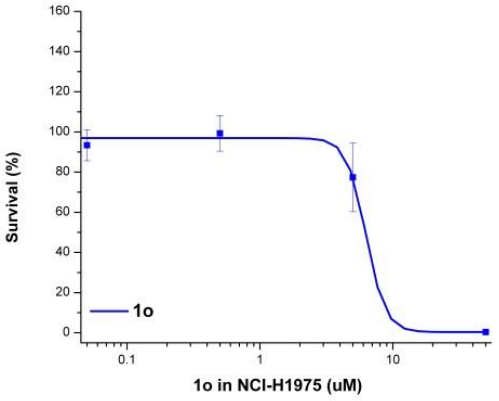
2f showed a consistent titration of metabolic activity, although this is not STAT3 specific as the titration was seen in both the MDA-MB-231 cells and A4 cells, indicating that **2f** efficiently causes altered metabolic activity through another mechanism of action to STAT3 SH-2 or DNA binding domain inhibition.

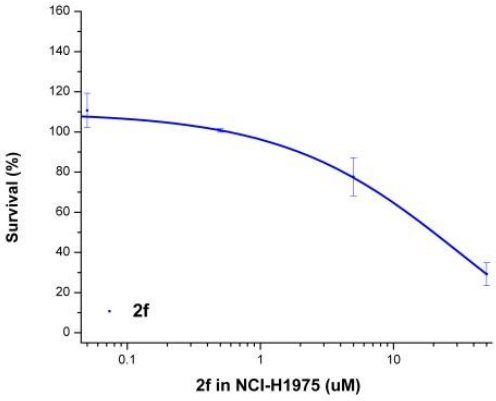
6.2.6 NCI-H1975 Non-Small Lung Carcinoma Study on Lead 2nd Generation Compounds

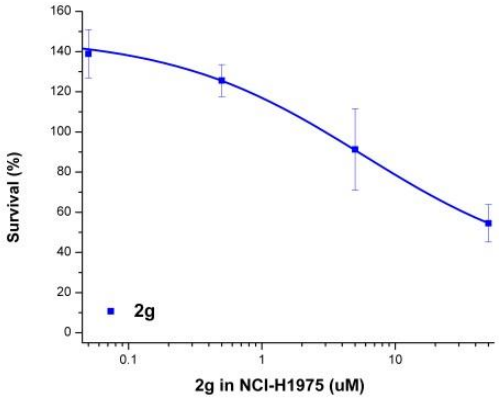
The six front running second generation compounds from the HTRF screen were tested for their cell cytotoxicity. An initial screen in the STAT3 dependent NCI-H1975 non-small lung carcinoma cell line was performed with an inhibitor range of 1 µM to 100 µM.

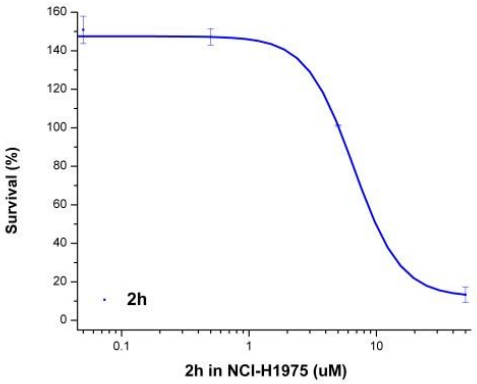












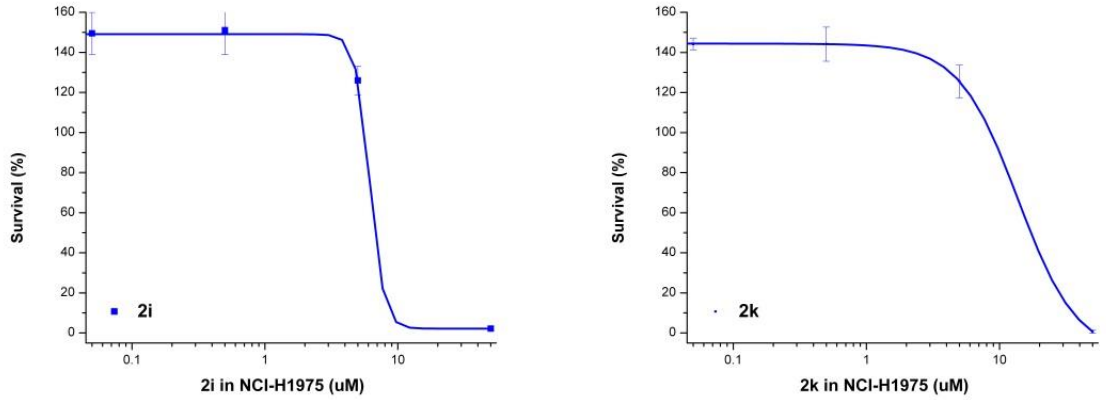


Figure 6.45 NCI-H1975 MTT cytotoxicity screen (as described in 2.2.6.5) in the presence of 2nd generation front runners. (Means derived from N=3 +/- SD)

Of the seven 2nd generation front runners **2f**, **2g** and **2h** did not achieve full cell death at 50 µM. The more targeted nature of the 2nd generation inhibitors ensures greater cell death at lower concentrations. A possible mechanism for the selective cytotoxicity for these compounds towards cancer cells could be due to the enhanced requirement for STAT3 activity in cancer cells, leading to a greater reduction in cell proliferation upon inhibition of STAT3 signaling by these compounds.

Modelling Studies

In silico modelling studies were conducted, as detailed in chapter 2, using STAT3 homodimer (PDB ID: 1BG1) and the 1st and 2nd generation of molecules. Modelling studies were carried out with the expertise of Dr Paul Jackson.

Table 6.22 shows the likelihood of binding to either the SH2 domain using the GRID and the GBSA score. An indication of whether that interaction is centrally based around the phosphorylated tyrosine residue (705) is also given. DNA-binding domain GRID and GBSA scores are indicated with an overall summary of whether the compound favours protein-protein or protein-DNA inhibition (P or D). The control STAT3 inhibitors Sta-21 and STATTIC were not modelled for reference as they are much smaller in terms of molecular weight (~200Da) than the test compounds. Binding energies would be incomparable as energies are always proportional to molecular weight. Small molecules will always have less favourable energies than larger molecules. The top sixteen most active molecules were ranked from the entire series and their position is shown here, if they featured. The greater the negativity of the GRID and GBSA scores, the stronger the binding affinity. The full prediction table can be seen in the appendices. Table 6.22 shows an abridged version of just the front running compounds from the HTRF and cytotoxicity assays.

+

Compound	GRID Score Protein (kcal/mol)	Y705 Binding Based on GRID	GBSA Score Protein (kcal/mol)	Y705 Binding Based on GBSA	Rank	GRID Score DNA (kcal/mol)	GBSA Score DNA (kcal/mol)	Preferentially binding Protein (P) or DNA (D)?
1a	-46.57	N	-51.96	N (Not SH-2)		-48.50	-50.25	P
2c	-53.40	N	-60.68	N		-49.44	-55.56	P
1c	-57.07	Y	-61.52	Y	16	Failed	Failed	P
1j	-58.18	Y	-62.50	Y	14	-31.95	-32.03	P
1k	-69.54	Y	-73.98	Y	2	-21.59	Failed	P
1n	-62.15	Y	-71.22	Y	10	-37.57	-53.68	P
1o	-65.74	Y	-63.04	Y	12	-27.31	-50.40	P
1r	-53.96	N	-57.33	N		-61.14	-68.20	D
1s	-59.78	N	-66.21	N		-52.55	-40.80	P
2e	-70.37	Y (and P704)	-70.47	Y (and P704)	5	-60.16	-70.56	P
2f	-62.78	Y (and P704)	-68.27	Y (and P704)	11	-71.10	-79.50	D
2g	-70.77	Y (and P704)	-76.39	Y(and P704)	1	-71.83	-80.30	P
2h	-71.78	Y(and P704)	-41.71	Y		-59.72	-64.97	P
2i	-70.26	Y(and P704)	-70.57	Y (and P704)	4	-61.59	-13.66	P
2k	-66.99	Y (and P704)	-67.14	Y (and P704)	8	-59.60	-67.45	P

Table 6.18 Docking scores of cytotoxically active front running compounds (both GRID and GBSA scoring in kcal/mol) in both the SH2 and DNA binding domains.

The front running compound (**2f**) was shown to bind preferentially to the DNA binding domain rather than the SH-2 domain and this was supported by this compounds performance in the cytotoxicity assays as 2f showed no specificity for either the MDA-MB-

231 or A4 cell lines. All the other front running compounds had a preference for the SH2 domain. The best compounds from cytotoxicity studies also featured in the top ten ranking of compounds (out of 45 original compounds) indicating a good correlation between experimental results and modelling predictions. The top 16 ranking table is shown in table 6.23 based on consistent binding affinity between GRID and GBSA scores.

Compound	Ranking
2g	1*
1k	2
2j	3
2i	4*
2e	5*
1h	6
1t	7
2k	8*
2d	9
1n	10*
2f	11*
1o	12
1f	13
1j	14
1m	15
1c	16

Table 6.19 Ranking of top 16 molecules according to a combined GRID and GBSA score (compound interferes with binding of both Y705 and P704)***

When both GRID and GBSA scoring factors are combined and ranked, the three most active STAT3 inhibitors from HTRF and MTT cytotoxicity assays **(2e, 2i** and **2k**) all appear in the top eight most interactive molecules based on docking calculations out of nearly fifty in the

series (Table 6.23). Furthermore, these molecules were all accommodated in a similar binding pocket, characterised by a number of hydrophobic (blue Figure 6.57), polar (red in Figure 6.57) and charged residues (green in Figure 6.57).

The STAT3 SH2 domain possesses three binding "hot spots": Firstly, the pTyr705-binding pocket with polar residues and contributing the largest binding enthalpy (two-thirds); secondly, the Leu706 subsite which is the most dynamic and hardest to target; and finally, a hydrophobic side pocket which is unique to STAT3 and very targetable (173). Many of the compounds in the series were seen to inhibit the binding of phosphorylated tyrosine residue 705 of homodimer B to the heptapeptide pocket (PYLKTKFI) of homodimer A (i.e. **2a**, **1b**, **2d**, **1c**, **1e**, **1f**, **1h**, **1j**, **1k**, **1m**, **1n**, **1o**, **2j**), and these results generally correlate well with observed HTRF and MTT assay results; these compounds are predominantly from the 1st generation series. What we observe with some of the 2nd generation compounds is an association of the molecule with proline 704 (P704), as well as Y705, and it is this greater stability that tends to correspond to greater potency in the HTRF assay and also greater specificity in the MTT cytotoxicity assay (**2i, 2k** and **2e**)., Therefore inhibition of both P704 and Y705 is seen to be crucial to higher specificity activity of STAT3 SH2 domain inhibitors.

This is exemplified in models of **2e** (Figure 6.58), which show the molecule snugly accommodated in a hydrophobic pocket (Val 637, Gly 618, Pro 639 and Pro 704), coupled with the formation of a single hydrogen bond between the carbonyl of the thiophene grouping and Ser 613. Furthermore, the shape-fit of the molecule is ideal for interaction in the PYLKTKFI binding pocket, evidenced by the high (>70 kcal/mol) GRID binding energy observed for the molecule.

6.2.6.1 Molecular Modelling of 2i

2i had a GRID score of -70.26 kcal/mol and a GBSA of -70.57 kcal/mol placing it 4th in the overall ranking. It was seen to interact with both P704 and Y705 and this result, along with its strong activity in HTRF (19% remaining activity at 40 µM) and a MDA-MB-231/A4 specificity ratio of 2.79, make this compound one of the most consistent and specific. Figure 6.57 shows the molecule (cyan) interacting with polar (red) residues of the STAT3 SH2 domain. Residues P704 and Y705 (magenta) of the second homodimer are seen to be inhibited from interacting with the same polar region.

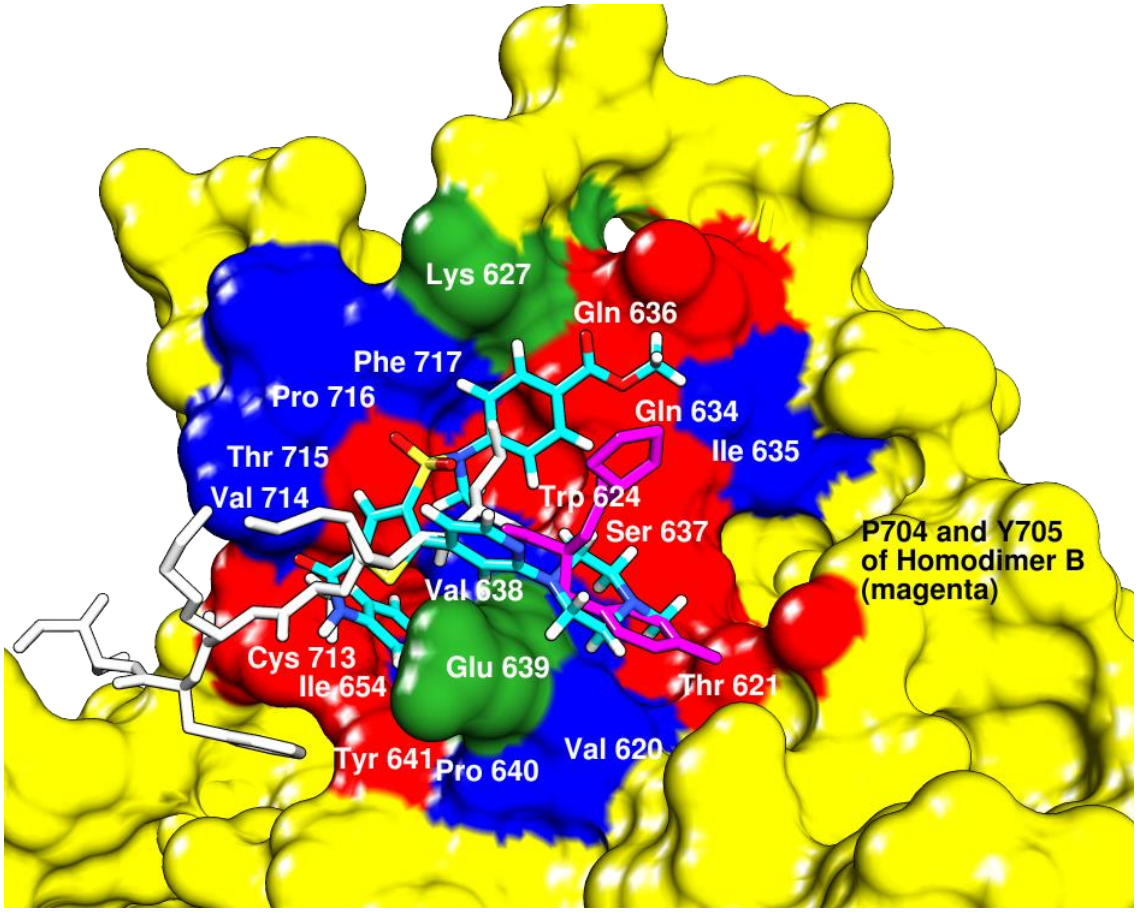


Figure 6.46 Molecular model of 2i (cyan) docked in the SH2 domain of the crystal structure of the STAT3 homodimer (PDB: 1BG1) showing the molecule inhibiting binding of P704 and Y705 (magenta) of homodimer B, along with much of the rest of the LKTKFI peptide of homodimer B (white). Prepared by Dr Paul Jackson

6.2.6.2 Molecular Modelling of 2e

2e had a GRID score of -70.37 kcal/mol and a GBSA of -70.47 kcal/mol placing it 5th in the overall ranking. It too has a strong affinity for both P704 and Y705. It was inactive in the HTRF assay which may be due to some inherent fluorescence properties of the R'$_1$ morpholine group. The activity of the molecule in the cytotoxicity assays was positive with an IC_{50} of 2.6 µM in MDA-MB-231 cells and no titration of inhibition in A4 cells, making **2e** an interesting molecule for further study and a front runner of this study (Figure 6.58).

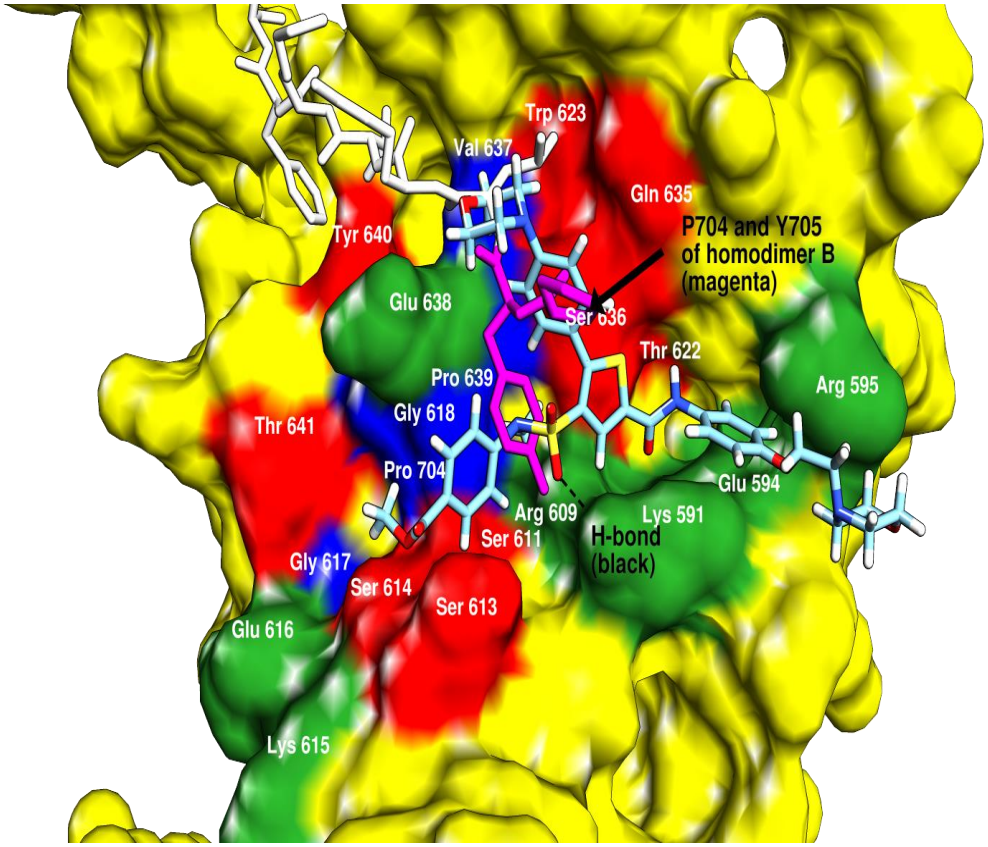


Figure 6.47 Molecular model of 2e (cyan) docked in the SH2 domain of the crystal structure of the STAT3 homodimer (PDB: 1BG1) showing the molecule inhibiting binding of P704 and Y705 (magenta) of homodimer B. The remainder of the octapeptide portion (LKTKFI) of homodimer B is shown in white. A single stabilising H-bond (black broken line) occurs between Ser 613 and the carbonyl group of the thiophene fragment. Prepared by Dr Paul Jackson

6.2.6.3 Molecular Modelling of 2k

2k had a GRID score of -66.99 kcal/mol and a GBSA of -67.14 kcal/mol placing it 8th in the overall ranking and characteristically with a strong affinity for both P704 and Y705. **2k** also had a HTRF activity of >100 and like **2e** this may be due to a non-specific reaction of the compound with the fluorophores. However, **2k**'s activity in the cytotoxicity assays was excellent, with an IC_{50} of 4.3 µM in MDA-MB-231 cells and of 27 µM in A4, giving **2k** a differential value of 6.18, which was the largest observed. **2k** therefore belongs in the lead molecule definition of this study (Figure 6.59).

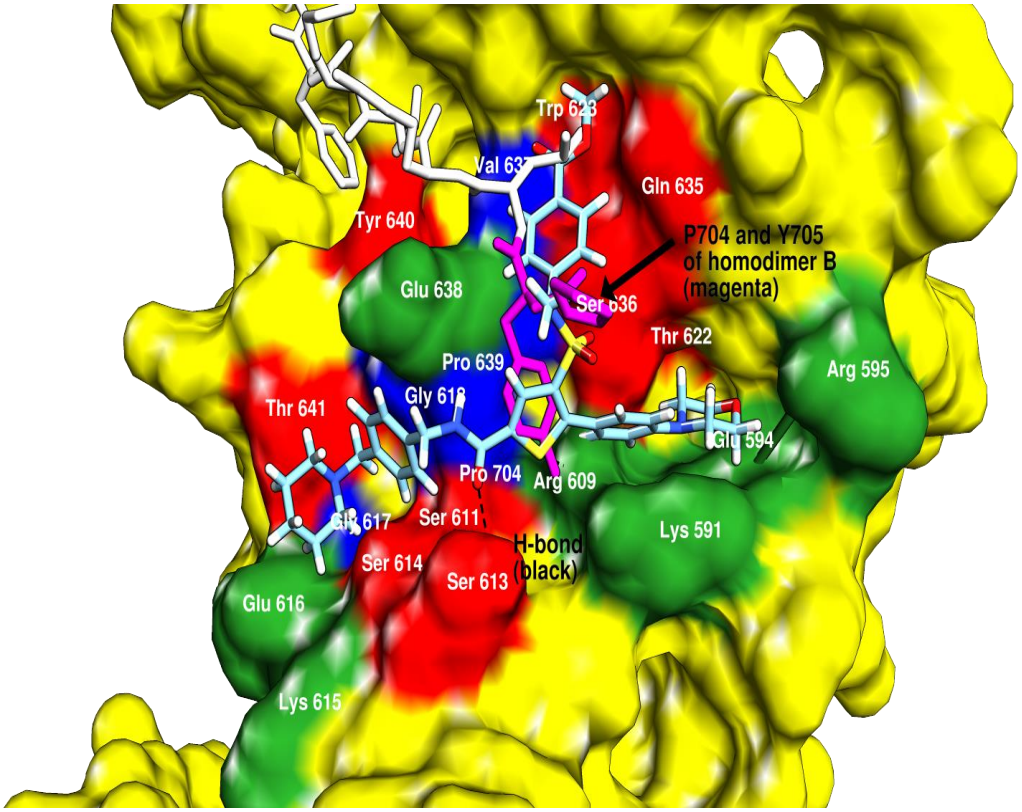


Figure 6.48 Molecular model of 2k (cyan) docked in the SH2 domain of the crystal structure of the STAT3 homodimer (PDB: 1BG1) showing the molecule inhibiting binding of P704 and Y705 (magenta) of homodimer B. The remainder of the octapeptide portion (LKTKFI) of homodimer B is shown in white. A single stabilising H-bond (black broken line) occurs between Ser 613 and the carbonyl group of the thiophene fragment, in a similar manner to 2e. Prepared by Dr Paul Jackson

6.2.6.4 Molecular Modelling of 2f

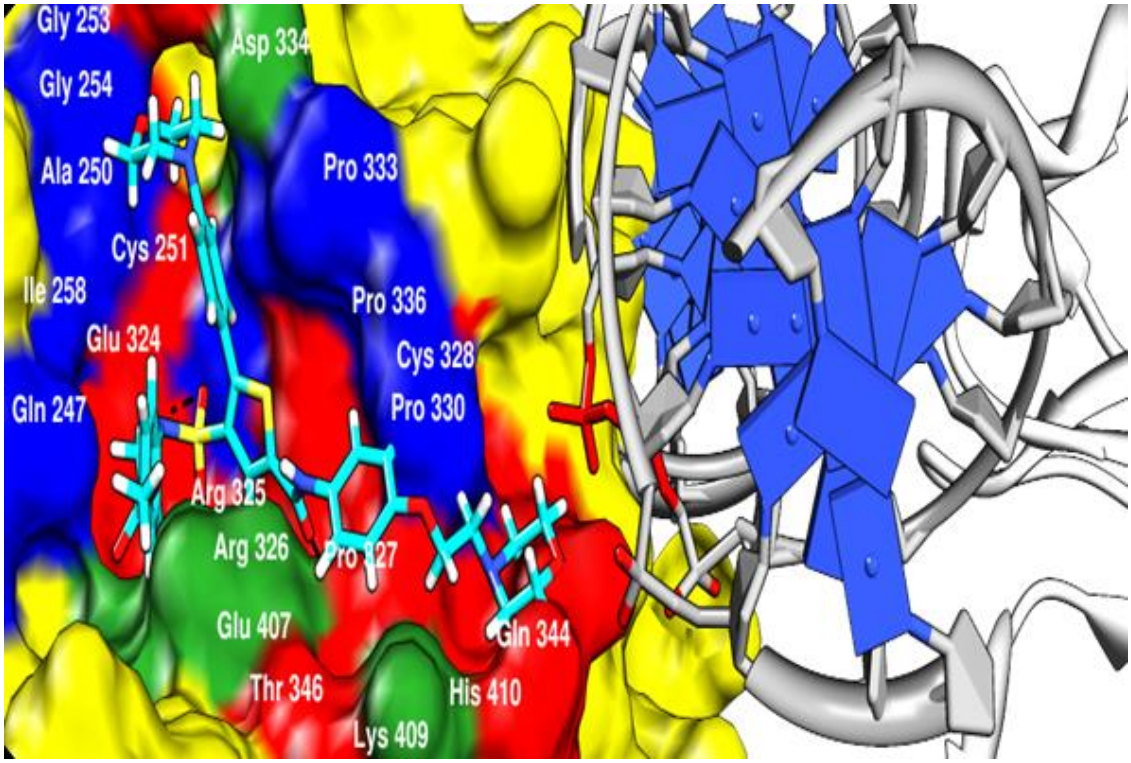


Figure 6.49 Molecular model of 2f (cyan) docked in the DNA binding domain of the crystal structure of the STAT3 homodimer (PDB: 1BG1) Prepared by Dr Paul Jackson.

Interestingly, a number of compounds designed for interaction with the SH2 domain were observed to possess a greater affinity for the DNA binding domain over the SH2 domain (Table 6.22). Examples of such compounds include **2a**, **1b**, **1r**, and **2f**. Of these molecules, **2f** exhibits potent cytotoxic activity against MDA-MB-231 cells (2.8 μM after 48 hrs incubation), suggesting a strong inhibitory effect on DNA-protein binding. This is reflected in docking results, where the molecule was observed to bind to the DNA-binding domain, potentially inhibiting the interaction of the STAT3 dimer with its consensus sequence (figure 6.60). The morpholine moiety is positioned ideally on the polar residues Glu 344 and His 410 to prevent interaction of the protein with its DNA consensus sequence.

Docking results correlate well with assay results (i.e. MTT and FRET binding assays) and support **2f**, **2g**, **1n**, **2e**, **2k** and **2i** as the most potent inhibitors of STAT3. This inhibition

occurs through a combination of amino acid-specific hydrogen bonding interactions, ideal shape-fit to the protein and inhibition of binding of P704 and Y705 of PYLKTKFI of homodimer B of the protein to its octapeptide binding pocket in homodimer A.

7 Conclusion and Future Work

The aim of this project was to initially develop an *in vitro* homogeneous time resolved FRET assay that could be used in low-, medium- and high-throughput modes for the discovery of novel STAT3 inhibitors. STAT3 is up-regulated and promotes the expression of several genes important in survival and metastasis, in many cancer types. There are currently no small-molecule STAT3 inhibitors in clinical use, and so there is a great need for the development of assays that can be used to screen molecules to identify lead compounds.

In order to develop an *in vitro* assay a reliable production and purification protocol of pure HIS-STAT3β_{TC} had to be developed. This was challenging and time-consuming as initial solubility and stability issues were encountered. However, experimental conditions were eventually established that allowed useful quantities (10 mg batches) of purified and stable protein to be obtained. Unphosphorylated and phosphorylated HIS-STAT3β_{TC} proteins were purified by initial ion exchange chromatography followed by Ni^{2+} affinity chromatography.

The purified recombinant proteins were filtered and concentration was determined by Bradford assay prior to their use in the developed homogenous time-resolved FRET (HTRF) assay. Development of this assay began with the optimisation of the Envision plate reader setup and calculation of the ΔF values of the buffer blanks and the two fluorophores blanks. A high and low calibration limit was established for the assay's boundaries.

A study of the variation between biotinylated and non-biotinylated M67 oligo was used to establish assay specificity, as no signal was seen when non-biotinylated DNA was used. The streptavidin-d2 fluorophore was unable to bind to the DNA consensus sequence and no FRET pair complex was formed. This assay also showed a correlation between the protein and DNA ratios. Similarly the difference between phosphorylated and un-phosphorylated STAT3 was equally pronounced; un-phosphorylated STAT3 had a very minimal signal up to 100 nM in the presence of both 20 and 40 nM biotinylated consensus DNA. The signal however went up proportionately when phosphorylated STAT3 was used. This was more

evidence of the need to form the whole complex (i.e. STAT3β_{TC} homodimer + anti-his europium and biotinylated M67 DNA + streptavidin-d2) as a base line of FRET signal that could then be disrupted with potential inhibitors.

STA-21 (an established inhibitor of STAT3 homodimerisation (148)) was used as a positive control to ensure that the inhibition of assay signal was proportionate to STA-21 concentration. SH-2 domain mimic peptides were also used to compete with native SH-2 domains to prevent homodimerisation of STAT3 and, therefore, HTRF assay signal.

The HTRF assay was then used to screen a library of 29 in-house produced potential inhibitors. This library was based on an initial hit compound RH-06 (159). The assay was used on inhibitor intermediates (i.e. scaffolds with a bromine residue in the R_2 position) and full compounds with unique R_1 and R_2 groups.

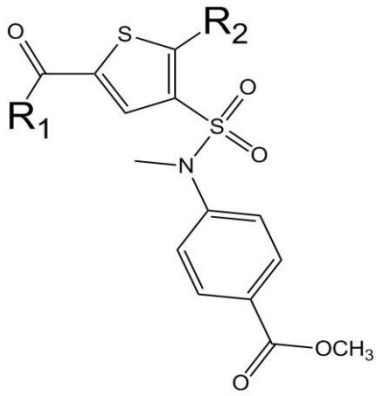


Figure 7.50 Basic scaffold of compound library

Point inhibition assays at 100 µM were carried out with all library compounds, and those compounds that showed a decrease in ΔF of greater than 50% compared with the DMSO control, were taken into a 10 µM point inhibition assay and titrated through a range of 250 nM to 25 µM.

The developed HTRF assay certainly has limitations in the consistency of data produced. For example, the DMSO only baseline signal would vary from run to run, and this made wider comparisons between individual experiments non-statistically valid. This may have been due to challenging experimental issues such as freeze/thaw reagent stability of both the fluorophores and protein. Compromises were also made regarding the plate reading equipment, such as the excitation and emission wavelengths, which were not theoretically optimal due to the available filters. However, when comparing inhibitors with each other within single experiments (in conjunction with the supporting cytotoxicity and modelling data), a clearer picture of the active functional R^1 and R^2 groups emerged.

The compound library was then tested against four cancer cell lines (HeLa, MDA-MB-231 breast cancer cells, NCI-H1975 lung cancer cells and A4 STAT3 null lung cancer cells) in order to deduce the compounds intercellular cytotoxicity using an MTT assay. STA-21 and STATTIC were tested in the cell lines as a positive control of cytotoxicity, and compared with results from the STAT3 dependent and null cell lines. Of the lead intermediate compounds from the HTRF assay, compound **1n** was the most cytotoxic in both HeLa and MDA-MB-231 cells. This cytotoxicity was STAT3 dependent as **1n** was more than five times as potent in STAT3 dependent cell lines.

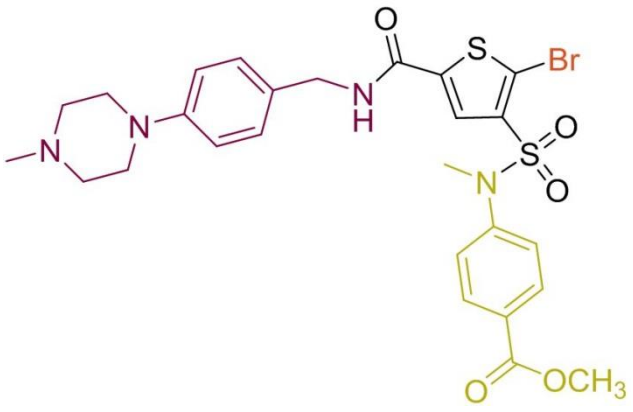


Figure 7.51 Structure of 1n

Four front running final compounds from the HTRF assay were tested for STAT3 dependent cytotoxicity. **2i** was the most potent and specific with an IC_{50} of 14.9 µM, which was nearly three times less than in the STAT3 null cell line.

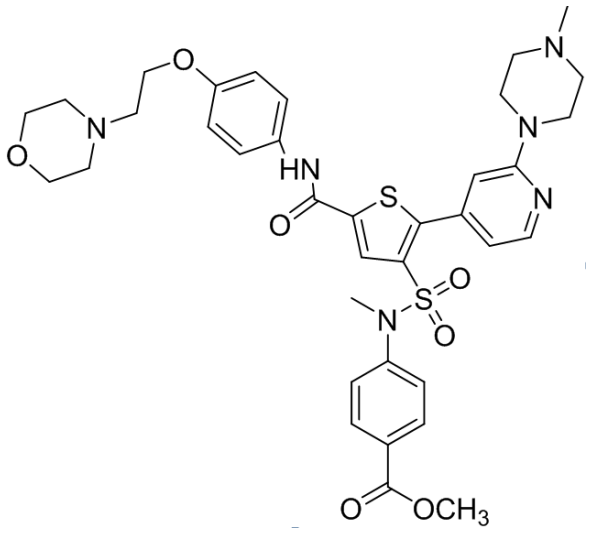


Figure 7.52 Structure of 2i

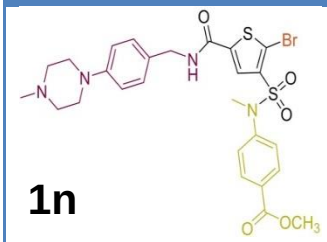
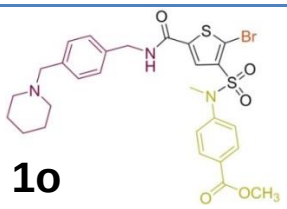
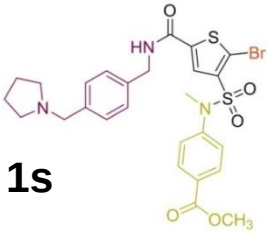
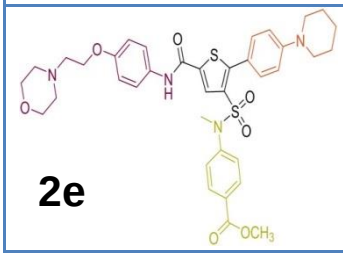
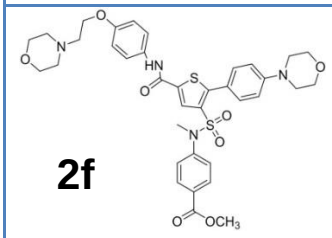
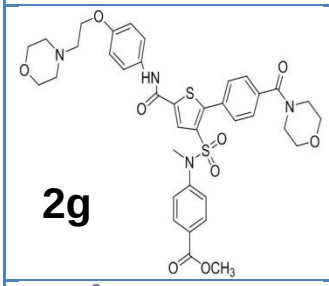
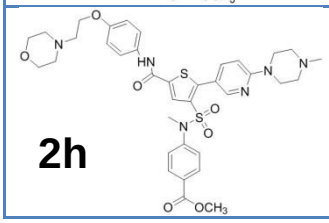
There were solubility limitations to the cytotoxicity assay as some compounds precipitated out of solution, even at low concentrations. Leading to the expected concentration of compound not being exposed to the cells, this could be seen in the well as precipitated crystals. These low solubility compounds may require a surfactant, such as Tween, in order to increase solubility and in future work a study of those otherwise promising compounds would be carried out. In further studies a MTS assay should be carried as an alternative and comparison to the MTT assay as the formazan formed from MTS is water-soluble and less toxic. Another cytotoxicity assay should also be carried out, such as a lactate dehydrogenase leakage (LDH) assay or a ATP-based method which is highly sensitive, in evaluating cell viability and proliferation. It is based on the quantitation of cellular ATP using the luciferin-luciferase reaction to produce bioluminescence. ATP degrades rapidly in dead cells and declines in injured cells. Therefore, the amount of ATP present in the cells is proportional to the number of viable cells in culture. This would allow a cytotoxic comparison

with MTT, (which has been shown to underestimate cytotoxicity(169)) and minimise false negative results.

Modelling studies were used to predict the likelihood of compounds binding to the Y705 area of the SH-2 domain and also the likelihood of binding to the DNA binding domain of STAT3.

The strength of binding affinities were ranked and compared with the assay data, and there was a good correlation between predicted and observed activity, with three compounds (**2e**, **2i** and **2k)** being the most active in HTRF and MTT cytotoxicity assay and also appearing in the top eight (out of 48) in docking calculations. In particular, the final scaffold structure containing three domains was seen to fit well into the Y705-P704 residue region of the SH-2 domain.

Summary table 7.56 combines the MDA-MB-231 and A4 IC_{50}s and looks at the differential ratio between the two. The table also collates the normalized ΔF percentage inhibition in HTRF assay compared with DMSO control and the ranking of the compound *in-silico* modelling studies. This list collates the 10 compounds of most interest for further future investigation. The criteria for selection in this table, out of the 42 compounds initially tested, was that they either were in the top 10 of the *in-silico* modelling study, they showed 50% inhibition in the HTRF assay or had a positive differential of 6 times for STAT3 vs STAT3 null cytotoxicity in the MTT assay. **2f** has been included in this table due to its interesting inhibition at the DNA binding domain of STAT3 as discussed in Section 6.2.6.4.

Compound structure	MDA-MB-231 IC_{50} (μM)	A4 IC_{50} (μM)	A4, MDA-MB-231 differential	HTRF Inhibition (%)	Modelling Rank
 1n	4.7	28.2	6.0	74	10
 1o	32.7	70.1	2.1	47	12
 1s	25.9	>100	3.8	75	16
 2e	2.6	>100	38.2	0	5
 2f	11.1	3.1	0.3	0	11
 2g	15	>100	6.7	56	1
 2h	40	2	0.1	53	>16

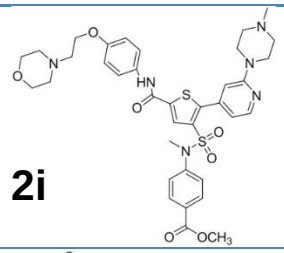
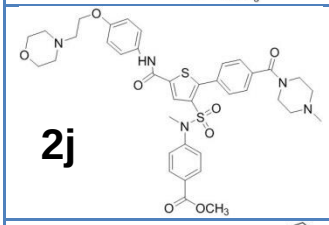
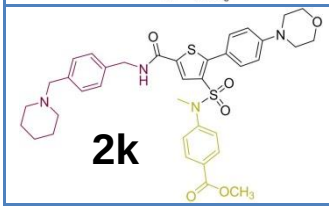
 2i	14.8	41.5	2.8	81	4
 2j	10	>100	10	0	3
 2k	4.4	27	6.2	0	8

Figure 7.53 Summary table of lead compound activity

Six of these molecules (2e, 2f, 2g, 2h, 2i and 2j) had an identical R_1 group, as shown in figure 7.61. This morpholine pharmacophore group is vital in forming a three pronged structure and it is this three pronged nature that allows the molecule to sit in the tyrosine 705 hot spot pocket as seen in the modelling projections. Earlier compounds (1a-1J, see page 124) had shorter, straighter R_1 which had neither the shape or molecule size to fit the SH-2 tyrosine 705 indentation. A key feature of the morpholine group is its ability to flexibly rotate around the carbon bridge which I propose allows rotation of the group whilst it is locating into its binding postion. Further analysis of this chemical feature is required in order to optimise binding. Further modelling is required in order to elucidate the morpholine's optimal rotation with the protein. This information could also suggest possible chemical enhancements that may be made to in the R_2 group position to further binding affinity.

Compound **2g** was the stand out molecule across the three measures, having the strongest binding affinity for the tyrosine 705 pocket out of all compounds. **2g** had a normalised ΔF inhibition of 56% in the HTRF assay, making it the fourth best compound overall in that

assay. **2g** also had a good IC_{50} in MDA-MB-231 cells of 15µM, however, it showed no cytotoxicity at all in the A4 cell line

The in silico modelling data consistently highlighted the molecules that performed well in HTRF and cytotoxicity assays, with all but one of the 10 most active compounds being predicted in the top 16 out of 42 compounds modelled. This suggests that the modelling screen is an effective method of testing for potential new R_1 and R_2 groups, prior to making the compound and testing it in the assays.

In conclusion, structure-based virtual screening has been utilised to research potential inhibitors of STAT3. Compounds of interest were able to inhibit STAT3 DNA-binding activity in a cell-free system and STAT3-driven cell viability in cancer cells,. Additionally, compound **1** antagonized STAT3 dimerization. Our molecular modeling analysis suggested that compound **1** might putatively function as an inhibitor of STAT3 dimerization. This study also validates the use of structure-based molecular docking to discover novel inhibitors of protein–protein interactions, which are typically considered difficult to target with small molecules. We envisage that these compound **1** may be employed as a useful scaffold for the development of more potent STAT3 dimerization inhibitors.

8 References

1. Bartek J, Lukas J. Are all cancer genes equal? Nature. 2001;411(6841):1001-2.
2. Weinberg RA. How cancer arises. Sci Am. 1996;275(3):62-70.
3. Hanahan D, Weinberg RA. The hallmarks of cancer. Cell. 100. United States2000. p. 57-70.
4. Kelland LR. Preclinical perspectives on platinum resistance. Drugs. 2000;59 Suppl 4:1-8; discussion 37-8.
5. Gibbs JB. Mechanism-based target identification and drug discovery in cancer research. Science. 287. United States2000. p. 1969-73.
6. Zanegina O, Aksianov E, Alexeevski AV, Karyagina A, Spirin S. Conserved features of complexes of TATA-box binding proteins with DNA. J Bioinform Comput Biol. 2016;14(2):1641007.
7. Terakawa T, Takada S. p53 dynamics upon response element recognition explored by molecular simulations. Sci Rep. 2015;5:17107.
8. Berg T. Inhibition of transcription factors with small organic molecules. Curr Opin Chem Biol. 2008;12(4):464-71.
9. Blume-Jensen P, Hunter T. Oncogenic kinase signalling. Nature. 2001;411(6835):355-65.
10. Pankratova EV, Stepchenko AG, Krylova ID, Portseva TN, Georgieva SG. Involvement of transcription factor Oct-1 in the regulation of JAK-STAT signaling pathway in cells of Burkitt lymphoma. Dokl Biochem Biophys. 2016;468(1):229-31.
11. Bild AH, Potti A, Nevins JR. Linking oncogenic pathways with therapeutic opportunities. Nat Rev Cancer. 6. England2006. p. 735-41.
12. Brivanlou AH, Darnell JE, Jr. Signal transduction and the control of gene expression. Science. 2002;295(5556):813-8.
13. Koehler AN. A complex task? Direct modulation of transcription factors with small molecules. Curr Opin Chem Biol. 2010;14(3):331-40.
14. Cheng AC, Coleman RG, Smyth KT, Cao Q, Soulard P, Caffrey DR, et al. Structure-based maximal affinity model predicts small-molecule druggability. Nat Biotechnol. 2007;25(1):71-5.
15. Liu BA, Jablonowski K, Raina M, Arcé M, Pawson T, Nash PD. The Human and Mouse Complement of SH2 Domain Proteins—Establishing the Boundaries of Phosphotyrosine Signaling. Molecular Cell. 2006;22(6):851-68.
16. Zhang X, Yue P, Fletcher S, Zhao W, Gunning PT, Turkson J. A novel small-molecule disrupts Stat3 SH2 domain-phosphotyrosine interactions and Stat3-dependent tumor processes. Biochemical Pharmacology. 2010;79(10):1398-409.
17. Arndt HD. Small molecule modulators of transcription. Angew Chem Int Ed Engl. 2006;45(28):4552-60.
18. Vogler M, Dinsdale D, Dyer MJ, Cohen GM. Bcl-2 inhibitors: small molecules with a big impact on cancer therapy. Cell Death Differ. 2009;16(3):360-7.
19. Lock R, Carol H, Houghton PJ, Morton CL, Kolb EA, Gorlick R, et al. Initial testing (stage 1) of the BH3 mimetic ABT-263 by the pediatric preclinical testing program. Pediatr Blood Cancer. 2008;50(6):1181-9.
20. Nakajima W, Sharma K, Hicks MA, Le N, Brown R, Krystal GW, et al. Combination with vorinostat overcomes ABT-263 (navitoclax) resistance of small cell lung cancer. Cancer Biol Ther. 2016;17(1):27-35.
21. Yang P, Chen W, Li X, Eilers G, He Q, Liu L, et al. Downregulation of cyclin D1 sensitizes cancer cells to MDM2 antagonist Nutlin-3. Oncotarget. 2016;7(22):32652-63.
22. Shin J-S, Ha J-H, He F, Muto Y, Ryu K-S, Yoon HS, et al. Structural insights into the dual-targeting mechanism of Nutlin-3. Biochemical and Biophysical Research Communications. 2012;420(1):48-53.
23. Hudson PJ, Souriau C. Engineered antibodies. Nat Med. 2003;9(1):129-34.

24. Vagner J, Qu H, Hruby VJ. Peptidomimetics, a synthetic tool of drug discovery. Curr Opin Chem Biol. 2008;12(3):292-6.
25. Wells JA, McClendon CL. Reaching for high-hanging fruit in drug discovery at protein-protein interfaces. Nature. 2007;450(7172):1001-9.
26. Copeland NG, Gilbert DJ, Schindler C, Zhong Z, Wen Z, Darnell JE, Jr., et al. Distribution of the mammalian Stat gene family in mouse chromosomes. Genomics. 29. United States1995. p. 225-8.
27. Horvath CM, Darnell JE. The state of the STATs: recent developments in the study of signal transduction to the nucleus. Curr Opin Cell Biol. 9. United States1997. p. 233-9.
28. Darnell JE, Kerr IM, Stark GR. Jak-Stat Pathways and Transcriptional Activation in Response to Ifns and Other Extracellular Signaling Proteins. Science. 1994;264(5164):1415-21.
29. Bromberg J. Stat proteins and oncogenesis. Journal of Clinical Investigation. 2002;109(9):1139-42.
30. Hou JZ, Schindler U, Henzel WJ, Ho TC, Brasseur M, Mcknight SL. An Interleukin-4-Induced Transcription Factor - Il-4 Stat. Science. 1994;265(5179):1701-6.
31. Xu XA, Sun YL, Hoey T. Cooperative DNA binding and sequence-selective recognition conferred by the STAT amino-terminal domain. Science. 1996;273(5276):794-7.
32. Vinkemeier U, Cohen SL, Moarefi I, Chait BT, Kuriyan J, Darnell JE. DNA binding of in vitro activated Stat1 alpha, Stat1 beta and truncated Stat1: Interaction between NH2-terminal domains stabilizes binding of two dimers to tandem DNA sites. Embo Journal. 1996;15(20):5616-26.
33. Horvath CM, Wen ZL, Darnell JE. A Stat Protein Domain That Determines DNA-Sequence Recognition Suggests a Novel DNA-Binding Domain. Gene Dev. 1995;9(8):984-94.
34. Schaefer TS, Sanders LK, Nathans D. Cooperative Transcriptional Activity of Jun and Stat3-Beta, a Short-Form of Stat3. P Natl Acad Sci USA. 1995;92(20):9097-101.
35. Huang Y, Qiu J, Dong S, Redell MS, Poli V, Mancini MA, et al. Stat3 isoforms, alpha and beta, demonstrate distinct intracellular dynamics with prolonged nuclear retention of Stat3beta mapping to its unique C-terminal end. J Biol Chem. 282. United States2007. p. 34958-67.
36. Schaefer TS, Sanders LK, Park OK, Nathans D. Functional differences between Stat3alpha and Stat3beta. Mol Cell Biol. 1997;17(9):5307-16.
37. Chakraborty A, White SM, Schaefer TS, Ball ED, Dyer KF, Tweardy DJ. Granulocyte colony-stimulating factor activation of Stat3 alpha and Stat3 beta in immature normal and leukemic human myeloid cells. Blood. 1996;88(7):2442-9.
38. Fang B. Genetic Interactions of STAT3 and Anticancer Drug Development. Cancers. 2014;6(1):494-525.
39. Hoey T, Zhang S, Schmidt N, Yu Q, Ramchandani S, Xu X, et al. Distinct requirements for the naturally occurring splice forms Stat4alpha and Stat4beta in IL-12 responses. EMBO J. 2003;22(16):4237-48.
40. Caldenhoven E, vanDijk TB, Solari R, Armstrong J, Raaijmakers JAM, Lammers JWJ, et al. STAT3 beta, a splice variant of transcription factor STAT3, is a dominant negative regulator of transcription. Journal of Biological Chemistry. 1996;271(22):13221-7.
41. Zhong Z, Wen ZL, Darnell JE. Stat3 - a Stat Family Member Activated by Tyrosine Phosphorylation in Response to Epidermal Growth-Factor and Interleukin-6. Science. 1994;264(5155):95-8.
42. Heinrich PC, Behrmann I, Muller-Newen G, Schaper F, Graeve L. Interleukin-6-type cytokine signalling through the gp130/Jak/STAT pathway. Biochem J. 1998;334 (Pt 2):297-314.
43. Hirano T, Ishihara K, Hibi M. Roles of STAT3 in mediating the cell growth, differentiation and survival signals relayed through the IL-6 family of cytokine receptors. Oncogene. 2000;19(21):2548-56.
44. Zhong Z, Wen ZL, Darnell JE. Stat3 and Stat4 - Members of the Family of Signal Transducers and Activators of Transcription. P Natl Acad Sci USA. 1994;91(11):4806-10.

45. Hemmann U, Gerhartz C, Heesel B, Sasse J, Kurapkat G, Grotzinger J, et al. Differential activation of acute phase response factor/Stat3 and Stat1 via the cytoplasmic domain of the interleukin 6 signal transducer gp130. II. Src homology SH2 domains define the specificity of stat factor activation. J Biol Chem. 1996;271(22):12999-3007.
46. Bromberg JF, Wrzeszczynska MH, Devgan G, Zhao YX, Pestell RG, Albanese C, et al. Stat3 as an oncogene. Cell. 1999;98(3):295-303.
47. Calo V, Migliavacca M, Bazan V, Macaluso M, Buscemi M, Gebbia N, et al. STAT proteins: from normal control of cellular events to tumorigenesis. J Cell Physiol. 2003;197(2):157-68.
48. Muller CW, Becker S, Groner B. Three-dimensional structure of the Stat3 beta homodimer bound to DNA. Nature. 1998;394(6689):145-51.
49. Chen XM, Vinkemeier U, Zhao YX, Jeruzalmi D, Darnell JE, Kuriyan J. Crystal structure of a tyrosine phosphorylated STAT-1 dimer bound to DNA. Cell. 1998;93(5):827-39.
50. Zhang T, Kee WH, Seow KT, Fung W, Cao X. The Coiled-Coil Domain of Stat3 Is Essential for Its SH2 Domain-Mediated Receptor Binding and Subsequent Activation Induced by Epidermal Growth Factor and Interleukin-6. Molecular and Cellular Biology. 2000;20(19):7132-9.
51. Yu Z, Kone BC. The STAT3 DNA-binding domain mediates interaction with NF-kappaB p65 and inducible nitric oxide synthase transrepression in mesangial cells. J Am Soc Nephrol. 2004;15(3):585-91.
52. Mertens C, Haripal B, Klinge S, Darnell JE. Mutations in the linker domain affect phospho-STAT3 function and suggest targets for interrupting STAT3 activity. Proc Natl Acad Sci U S A. 2015;112(48):14811-6.
53. Becker S, Groner B, Muller CW. Three-dimensional structure of the Stat3beta homodimer bound to DNA. Nature. 1998;394(6689):145-51.
54. Kishimoto T. Interleukin-6: from basic science to medicine--40 years in immunology. Annu Rev Immunol. 2005;23:1-21.
55. Tamiya T, Kashiwagi I, Takahashi R, Yasukawa H, Yoshimura A. Suppressors of cytokine signaling (SOCS) proteins and JAK/STAT pathways: regulation of T-cell inflammation by SOCS1 and SOCS3. Arterioscler Thromb Vasc Biol. 2011;31(5):980-5.
56. Sasaki A, Yasukawa H, Suzuki A, Kamizono S, Syoda T, Kinjyo I, et al. Cytokine-inducible SH2 protein-3 (CIS3/SOCS3) inhibits Janus tyrosine kinase by binding through the N-terminal kinase inhibitory region as well as SH2 domain. Genes Cells. 1999;4(6):339-51.
57. Yan C, Yu C, Zhang D, Cui Y, Zhou J, Cui S. Protein Inhibitor of Activated STAT Y (PIASy) Regulates Insulin Secretion by Interacting with LIM Homeodomain Transcription Factor Isl1. Sci Rep. 2016;6:39308.
58. Chung CD, Liao J, Liu B, Rao X, Jay P, Berta P, et al. Specific inhibition of Stat3 signal transduction by PIAS3. Science. 1997;278(5344):1803-5.
59. Lee J, Beliakoff J, Sun Z. The novel PIAS-like protein hZimp10 is a transcriptional co-activator of the p53 tumor suppressor. Nucleic Acids Res. 2007;35(13):4523-34.
60. Nie Y, Erion DM, Yuan Z, Dietrich M, Shulman GI, Horvath TL, et al. STAT3 inhibition of gluconeogenesis is downregulated by SirT1. Nat Cell Biol. 11. England2009. p. 492-500.
61. Barillas-Mury C, Han YS, Seeley D, Kafatos FC. Anopheles gambiae Ag-STAT, a new insect member of the STAT family, is activated in response to bacterial infection. EMBO J. 1999;18(4):959-67.
62. Braun DA, Fribourg M, Sealfon SC. Cytokine response is determined by duration of receptor and signal transducers and activators of transcription 3 (STAT3) activation. J Biol Chem. 2013;288(5):2986-93.
63. Gallucci RM, Simeonova PP, Matheson JM, Kommineni C, Guriel JL, Sugawara T, et al. Impaired cutaneous wound healing in interleukin-6-deficient and immunosuppressed mice. Faseb j. 2000;14(15):2525-31.
64. Zong CS, Chan J, Levy DE, Horvath C, Sadowski HB, Wang LH. Mechanism of STAT3 activation by insulin-like growth factor I receptor. J Biol Chem. 2000;275(20):15099-105.

65. Carvalheira JB, Siloto RM, Ignacchitti I, Brenelli SL, Carvalho CR, Leite A, et al. Insulin modulates leptin-induced STAT3 activation in rat hypothalamus. FEBS Lett. 2001;500(3):119-24.
66. Takeda K, Noguchi K, Shi W, Tanaka T, Matsumoto M, Yoshida N, et al. Targeted disruption of the mouse Stat3 gene leads to early embryonic lethality. Proc Natl Acad Sci U S A. 1997;94(8):3801-4.
67. Niso-Santano M, Shen S, Adjemian S, Malik SA, Marino G, Lachkar S, et al. Direct interaction between STAT3 and EIF2AK2 controls fatty acid-induced autophagy. Autophagy. 2013;9(3):415-7.
68. Sano S, Takahama Y, Sugawara T, Kosaka H, Itami S, Yoshikawa K, et al. Stat3 in thymic epithelial cells is essential for postnatal maintenance of thymic architecture and thymocyte survival. Immunity. 2001;15(2):261-73.
69. Chapman RS, Lourenco PC, Tonner E, Flint DJ, Selbert S, Takeda K, et al. Suppression of epithelial apoptosis and delayed mammary gland involution in mice with a conditional knockout of Stat3. Genes Dev. 1999;13(19):2604-16.
70. Niu G, Wright KL, Ma Y, Wright GM, Huang M, Irby R, et al. Role of Stat3 in Regulating p53 Expression and Function. Molecular and Cellular Biology. 2005;25(17):7432-40.
71. Yu H, Jove R. The STATs of cancer--new molecular targets come of age. Nat Rev Cancer. 2004;4(2):97-105.
72. Carpenter RL, Lo H-W. STAT3 Target Genes Relevant to Human Cancers. Cancers. 2014;6(2):897-925.
73. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2(5):401-4.
74. Jones AV, Kreil S, Zoi K, Waghorn K, Curtis C, Zhang L, et al. Widespread occurrence of the JAK2 V617F mutation in chronic myeloproliferative disorders. Blood. 2005;106(6):2162-8.
75. Fleming TP, Saxena A, Clark WC, Robertson JT, Oldfield EH, Aaronson SA, et al. Amplification and/or overexpression of platelet-derived growth factor receptors and epidermal growth factor receptor in human glial tumors. Cancer Res. 1992;52(16):4550-3.
76. Gao SP, Mark KG, Leslie K, Pao W, Motoi N, Gerald WL, et al. Mutations in the EGFR kinase domain mediate STAT3 activation via IL-6 production in human lung adenocarcinomas. J Clin Invest. 2007;117(12):3846-56.
77. Haura EB, Zheng Z, Song L, Cantor A, Bepler G. Activated epidermal growth factor receptor-Stat-3 signaling promotes tumor survival in vivo in non-small cell lung cancer. Clin Cancer Res. 11. United States2005. p. 8288-94.
78. Veeriah S, Brennan C, Meng S, Singh B, Fagin JA, Solit DB, et al. The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. Proc Natl Acad Sci U S A. 2009;106(23):9435-40.
79. Ewen ME, Lamb J. The activities of cyclin D1 that drive tumorigenesis. Trends Mol Med. 2004;10(4):158-62.
80. Bowman T, Broome MA, Sinibaldi D, Wharton W, Pledger WJ, Sedivy JM, et al. Stat3-mediated Myc expression is required for Src transformation and PDGF-induced mitogenesis. Proc Natl Acad Sci U S A. 2001;98(13):7319-24.
81. Zhuang L, Lee CS, Scolyer RA, McCarthy SW, Zhang XD, Thompson JF, et al. Mcl-1, Bcl-XL and Stat3 expression are associated with progression of melanoma whereas Bcl-2, AP-2 and MITF levels decrease during progression of melanoma. Mod Pathol. 2007;20(4):416-26.
82. Bowman T, Garcia R, Turkson J, Jove R. STATs in oncogenesis. Oncogene. 2000;19(21):2474-88.
83. Otake Y, Soundararajan S, Sengupta TK, Kio EA, Smith JC, Pineda-Roman M, et al. Overexpression of nucleolin in chronic lymphocytic leukemia cells induces stabilization of bcl2 mRNA. Blood. 2007;109(7):3069-75.

84. Epling-Burnette PK, Liu JH, Catlett-Falcone R, Turkson J, Oshiro M, Kothapalli R, et al. Inhibition of STAT3 signaling leads to apoptosis of leukemic large granular lymphocytes and decreased Mcl-1 expression. J Clin Invest. 2001;107(3):351-62.
85. Xie TX, Huang FJ, Aldape KD, Kang SH, Liu M, Gershenwald JE, et al. Activation of stat3 in human melanoma promotes brain metastasis. Cancer Res. 2006;66(6):3188-96.
86. Zhang F, Wang Z, Fan Y, Xu Q, Ji W, Tian R, et al. Elevated STAT3 Signaling-Mediated Upregulation of MMP-2/9 Confers Enhanced Invasion Ability in Multidrug-Resistant Breast Cancer Cells. Int J Mol Sci. 2015;16(10):24772-90.
87. Wang Z, Zhu S, Shen M, Liu J, Wang M, Li C, et al. STAT3 is involved in esophageal carcinogenesis through regulation of Oct-1. Carcinogenesis. 2013;34(3):678-88.
88. Martin-Villar E, Borda-d'Agua B, Carrasco-Ramirez P, Renart J, Parsons M, Quintanilla M, et al. Podoplanin mediates ECM degradation by squamous carcinoma cells through control of invadopodia stability. Oncogene. 2015;34(34):4531-44.
89. Fujii M, Honma M, Takahashi H, Ishida-Yamamoto A, Iizuka H. Intercellular contact augments epidermal growth factor receptor (EGFR) and signal transducer and activator of transcription 3 (STAT3)-activation which increases podoplanin-expression in order to promote squamous cell carcinoma motility. Cell Signal. 2013;25(4):760-5.
90. Niu GL, Wright KL, Huang M, Song LX, Haura E, Turkson J, et al. Constitutive Stat3 activity up-regulates VEGF expression and tumor angiogenesis. Oncogene. 2002;21(13):2000-8.
91. Alonzi T, Middleton G, Wyatt S, Buchman V, Betz UA, Muller W, et al. Role of STAT3 and PI 3-kinase/Akt in mediating the survival actions of cytokines on sensory neurons. Mol Cell Neurosci. 18. United States: 2001 Academic Press.; 2001. p. 270-82.
92. Chen Z, Han ZC. STAT3: a critical transcription activator in angiogenesis. Med Res Rev. 2008;28(2):185-200.
93. Carballo M, Conde M, El Bekay R, Martin-Nieto J, Camacho MJ, Monteseirin J, et al. Oxidative stress triggers STAT3 tyrosine phosphorylation and nuclear translocation in human lymphocytes. J Biol Chem. 1999;274(25):17580-6.
94. Ihle JN. The Stat family in cytokine signaling. Curr Opin Cell Biol. 13. United States2001. p. 211-7.
95. Schindler C, Strehlow I. Cytokines and STAT signaling. Adv Pharmacol. 2000;47:113-74.
96. Liongue C, Ward AC. Evolution of the JAK-STAT pathway. JAK-STAT. 2013;2(1):e22756.
97. Turkson J, Kim JS, Zhang SM, Yuan J, Huang M, Glenn M, et al. Novel peptidomimetic inhibitors of signal transducer and activator of transcription 3 dimerization and biological activity. Molecular Cancer Therapeutics. 2004;3(3):261-9.
98. Greenlund AC, Morales MO, Viviano BL, Yan H, Krolewski J, Schreiber RD. Stat recruitment by tyrosine-phosphorylated cytokine receptors: an ordered reversible affinity-driven process. Immunity. 2. United States1995. p. 677-87.
99. Lillemeier BF, Koster M, Kerr IM. STAT1 from the cell membrane to the DNA. Embo j. 2001;20(10):2508-17.
100. Bhattacharya S, Schindler C. Regulation of Stat3 nuclear export. J Clin Invest. 2003;111(4):553-9.
101. Heim MH. The Jak-STAT pathway: cytokine signalling from the receptor to the nucleus. J Recept Signal Transduct Res. 1999;19(1-4):75-120.
102. Cao X, Tay A, Guy GR, Tan YH. Activation and association of Stat3 with Src in v-Src-transformed cell lines. Mol Cell Biol. 1996;16(4):1595-603.
103. Marrero MB, Schieffer B, Paxton WG, Heerdt L, Berk BC, Delafontaine P, et al. Direct stimulation of Jak/STAT pathway by the angiotensin II AT1 receptor. Nature. 1995;375(6528):247-50.
104. Leonard WJ, O'Shea JJ. Jaks and STATs: biological implications. Annu Rev Immunol. 1998;16:293-322.

105. Nkansah E, Shah R, Collie GW, Parkinson GN, Palmer J, Rahman KM, et al. Observation of unphosphorylated STAT3 core protein binding to target dsDNA by PEMSA and X-ray crystallography. FEBS Lett. 2013;587(7):833-9.
106. Wen Z, Zhong Z, Darnell JE, Jr. Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. Cell. 1995;82(2):241-50.
107. Decker T, Kovarik P. Serine phosphorylation of STATs. Oncogene. 2000;19(21):2628-37.
108. Kovarik P, Mangold M, Ramsauer K, Heidari H, Steinborn R, Zotter A, et al. Specificity of signaling by STAT1 depends on SH2 and C-terminal domains that regulate Ser727 phosphorylation, differentially affecting specific target gene expression. EMBO J. 2001;20(1-2):91-100.
109. Kovarik P, Stoiber D, Eyers PA, Menghini R, Neininger A, Gaestel M, et al. Stress-induced phosphorylation of STAT1 at Ser727 requires p38 mitogen-activated protein kinase whereas IFN-gamma uses a different signaling pathway. Proc Natl Acad Sci U S A. 1999;96(24):13956-61.
110. Leaman DW, Leung S, Li X, Stark GR. Regulation of STAT-dependent pathways by growth factors and cytokines. FASEB J. 1996;10(14):1578-88.
111. Collins AS, McCoy CE, Lloyd AT, O'Farrelly C, Stevenson NJ. miR-19a: an effective regulator of SOCS3 and enhancer of JAK-STAT signalling. PLoS One. 2013;8(7):e69090.
112. Zhang P, Yang B, Yao YY, Zhong LX, Chen XY, Kong QY, et al. PIAS3, SHP2 and SOCS3 Expression patterns in Cervical Cancers: Relevance with activation and resveratrol-caused inactivation of STAT3 signaling. Gynecol Oncol. 2015;139(3):529-35.
113. Groner B, Lucks P, Borghouts C. The function of Stat3 in tumor cells and their microenvironment. Semin Cell Dev Biol. 19. England2008. p. 341-50.
114. Letellier E, Haan S. SOCS2: physiological and pathological functions. Front Biosci (Elite Ed). 2016;8:189-204.
115. Greenhalgh CJ, Bertolino P, Asa SL, Metcalf D, Corbin JE, Adams TE, et al. Growth enhancement in suppressor of cytokine signaling 2 (SOCS-2)-deficient mice is dependent on signal transducer and activator of transcription 5b (STAT5b). Mol Endocrinol. 2002;16(6):1394-406.
116. Li J, Cui J, Zhang J, Liu Y, Han L, Jia C, et al. PIAS3, an inhibitor of STAT3, has intensively negative association with the survival of gastric cancer. Int J Clin Exp Med. 2015;8(1):682-9.
117. Hendry L, John S. Regulation of STAT signalling by proteolytic processing. Eur J Biochem. 271. Germany2004. p. 4613-20.
118. Gupta S, Yan H, Wong LH, Ralph S, Krolewski J, Schindler C. The SH2 domains of Stat1 and Stat2 mediate multiple interactions in the transduction of IFN-alpha signals. EMBO J. 1996;15(5):1075-84.
119. Pawson T, Gish GD, Nash P. SH2 domains, interaction modules and cellular wiring. Trends Cell Biol. 2001;11(12):504-11.
120. Kuriyan J, Cowburn D. Modular peptide recognition domains in eukaryotic signaling. Annu Rev Biophys Biomol Struct. 1997;26:259-88.
121. Begitt A, Meyer T, van Rossum M, Vinkemeier U. Nucleocytoplasmic translocation of Stat1 is regulated by a leucine-rich export signal in the coiled-coil domain. Proc Natl Acad Sci U S A. 97. United States2000. p. 10418-23.
122. Decker T, Kovarik P, Meinke A. GAS elements: a few nucleotides with a major impact on cytokine-induced gene expression. J Interferon Cytokine Res. 1997;17(3):121-34.
123. Kawata T, Shevchenko A, Fukuzawa M, Jermyn KA, Totty NF, Zhukovskaya NV, et al. SH2 signaling in a lower eukaryote: a STAT protein that regulates stalk cell differentiation in dictyostelium. Cell. 89. United States1997. p. 909-16.
124. Ehret GB, Reichenbach P, Schindler U, Horvath CM, Fritz S, Nabholz M, et al. DNA binding specificity of different STAT proteins. Comparison of in vitro specificity with natural target sites. J Biol Chem. 2001;276(9):6675-88.
125. Deng J, Grande F, Neamati N. Small molecule inhibitors of Stat3 signaling pathway. Curr Cancer Drug Targets. 2007;7(1):91-107.

126. Germain D, Frank DA. Targeting the cytoplasmic and nuclear functions of signal transducers and activators of transcription 3 for cancer therapy. Clin Cancer Res. 13. United States2007. p. 5665-9.
127. Haan S, Kortylewski M, Behrmann I, Muller-Esterl W, Heinrich PC, Schaper F. Cytoplasmic STAT proteins associate prior to activation. Biochem J. 2000;345 Pt 3:417-21.
128. Turkson J, Ryan D, Kim JS, Zhang Y, Chen Z, Haura E, et al. Phosphotyrosyl peptides block Stat3-mediated DNA binding activity, gene regulation, and cell transformation. Journal of Biological Chemistry. 2001;276(48):45443-55.
129. Gunning PT, Katt WP, Glenn M, Siddiquee K, Kim JS, Jove R, et al. Isoform selective inhibition of STAT1 or STAT3 homo-dimerization via peptidomimetic probes: structural recognition of STAT SH2 domains. Bioorg Med Chem Lett. 17. England2007. p. 1875-8.
130. Zhang X, Yue P, Page BD, Li T, Zhao W, Namanja AT, et al. Orally bioavailable small-molecule inhibitor of transcription factor Stat3 regresses human breast and lung cancer xenografts. Proc Natl Acad Sci U S A. 2012;109(24):9623-8.
131. McMurray JS. A new small-molecule Stat3 inhibitor. Chem Biol. 2006;13(11):1123-4.
132. Coleman DR, Ren ZY, Mandal PK, Cameron AG, Dyer GA, Muranjan S, et al. Investigation of the binding determinants of phosphopeptides targeted to the Src homology 2 domain of the signal transducer and activator of transcription 3. Development of a high-affinity peptide inhibitor. Journal of Medicinal Chemistry. 2005;48(21):6661-70.
133. Ren ZY, Cabell LA, Schaefer TS, McMurray JS. Identification of a high-affinity phosphopeptide inhibitor of Stat3. Bioorganic & Medicinal Chemistry Letters. 2003;13(4):633-6.
134. Dourlat J, Liu WQ, Sancier F, Edmonds T, Pamonsinlapatham P, Cruzalegui F, et al. A novel non-phosphorylated potential antitumoral peptide inhibits STAT3 biological activity. Biochimie. 91. France2009. p. 996-1002.
135. Lai PS, Rosa DA, Magdy Ali A, Gomez-Biagi RF, Ball DP, Shouksmith AE, et al. A STAT inhibitor patent review: progress since 2011. Expert Opin Ther Pat. 2015;25(12):1397-421.
136. Schust J, Sperl B, Hollis A, Mayer TU, Berg T. Stattic: A small-molecule inhibitor of STAT3 activation and dimerization. Chem Biol. 2006;13(11):1235-42.
137. Hao W, Hu Y, Niu C, Huang X, Chang CP, Gibbons J, et al. Discovery of the catechol structural moiety as a Stat3 SH2 domain inhibitor by virtual screening. Bioorg Med Chem Lett. 2008;18(18):4988-92.
138. Siddiquee K, Zhang S, Guida WC, Blaskovich MA, Greedy B, Lawrence HR, et al. Selective chemical probe inhibitor of Stat3, identified through structure-based virtual screening, induces antitumor activity. Proceedings of the National Academy of Sciences of the United States of America. 2007;104(18):7391-6.
139. Zhang XL, Yue PB, Page BDG, Li TS, Zhao W, Namanja AT, et al. Orally bioavailable small-molecule inhibitor of transcription factor Stat3 regresses human breast and lung cancer xenografts. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(24):9623-8.
140. Urlam MK, Pireddu R, Ge Y, Zhang X, Sun Y, Lawrence HR, et al. Development of new N-Arylbenzamides as STAT3 Dimerization Inhibitors. Medchemcomm. 2013;4(6):932-41.
141. Xu X, Kasembeli MM, Jiang X, Tweardy BJ, Tweardy DJ. Chemical probes that competitively and selectively inhibit Stat3 activation. PLoS One. 2009;4(3):e4783.
142. Dave B, Landis MD, Tweardy DJ, Chang JC, Dobrolecki LE, Wu MF, et al. Selective small molecule Stat3 inhibitor reduces breast cancer tumor-initiating cells and improves recurrence free survival in a human-xenograft model. PLoS One. 2012;7(8):e30207.
143. Shin DS, Kim HN, Shin KD, Yoon YJ, Kim SJ, Han DC, et al. Cryptotanshinone inhibits constitutive signal transducer and activator of transcription 3 function through blocking the dimerization in DU145 prostate cancer cells. Cancer Res. 2009;69(1):193-202.
144. Pan J, Lee Y, Zhang Q, Xiong D, Wan TC, Wang Y, et al. Honokiol Decreases Lung Cancer Metastasis through Inhibition of the STAT3 Signaling Pathway. Cancer Prev Res (Phila). 2016.

145. Mantaj J, Rahman SM, Bokshi B, Hasan CM, Jackson PJ, Parsons RB, et al. Crispene E, a cis-clerodane diterpene inhibits STAT3 dimerization in breast cancer cells. Org Biomol Chem. 2015;13(13):3882-6.
146. Song H, Wang RX, Wang SM, Lin J. A low-molecular-weight compound discovered through virtual database screening inhibits Stat3 function in breast cancer cells. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(13):4700-5.
147. Ahmad SF, Ansari MA, Nadeem A, Zoheir KMA, Bakheet SA, Alsaad AMS, et al. STA-21, a STAT-3 inhibitor, attenuates the development and progression of inflammation in collagen antibody-induced arthritis. Immunobiology.
148. Park JS, Kwok SK, Lim MA, Kim EK, Ryu JG, Kim SM, et al. STA-21, a promising STAT-3 inhibitor that reciprocally regulates Th17 and Treg cells, inhibits osteoclastogenesis in mice and humans and alleviates autoimmune inflammation in an experimental model of rheumatoid arthritis. Arthritis Rheumatol. 2014;66(4):918-29.
149. Fuh B, Sobo M, Cen L, Josiah D, Hutzen B, Cisek K, et al. LLL-3 inhibits STAT3 activity, suppresses glioblastoma cell growth and prolongs survival in a mouse glioblastoma model. Br J Cancer. 2009;100(1):106-12.
150. Sen M, Tosca PJ, Zwayer C, Ryan MJ, Johnson JD, Knostman KAB, et al. Lack of toxicity of a STAT3 decoy oligonucleotide. Cancer Chemoth Pharm. 2009;63(6):983-95.
151. Sen M, Thomas SM, Kim S, Yeh JI, Ferris RL, Johnson JT, et al. First-in-Human Trial of a STAT3 Decoy Oligonucleotide in Head and Neck Tumors: Implications for Cancer Therapy. Cancer Discov. 2012;2(8):694-705.
152. Shen J, Li R, Li G. Inhibitory Effects of Decoy-ODN Targeting Activated STAT3 on Human Glioma Growth In Vivo. In Vivo. 2009;23(2):237-43.
153. Souissi I, Ladam P, Cognet JAH, Le Coquil S, Varin-Blank N, Baran-Marszak F, et al. A STAT3-inhibitory hairpin decoy oligodeoxynucleotide discriminates between STAT1 and STAT3 and induces death in a human colon carcinoma cell line. Molecular Cancer. 2012;11.
154. Huang W, Dong Z, Chen Y, Wang F, Wang CJ, Peng H, et al. Small-molecule inhibitors targeting the DNA-binding domain of STAT3 suppress tumor growth, metastasis and STAT3 target gene expression in vivo. Oncogene. 35. England2016. p. 802.
155. Grandis JR, Drenning SD, Zeng Q, Watkins SC, Melhem MF, Endo S, et al. Constitutive activation of Stat3 signaling abrogates apoptosis in squamous cell carcinogenesis in vivo. Proc Natl Acad Sci U S A. 97. United States2000. p. 4227-32.
156. Barton BE, Murphy TF, Shu P, Huang HF, Meyenhofen M, Barton A. Novel single-stranded oligonucleotides that inhibit signal transducer and activator of transcription 3 induce apoptosis in vitro and in vivo in prostate cancer cell lines. Molecular Cancer Therapeutics. 2004;3(10):1183-91.
157. Li WC, Ye SL, Sun RX, Liu YK, Tang ZY, Kim Y, et al. Inhibition of growth and metastasis of human hepatocellular carcinoma by antisense oligonucleotide targeting signal transducer and activator of transcription 3. Clinical Cancer Research. 2006;12(23):7140-8.
158. Burel SA, Han SR, Lee HS, Norris DA, Lee BS, Machemer T, et al. Preclinical Evaluation of the Toxicological Effects of a Novel Constrained Ethyl Modified Antisense Compound Targeting Signal Transducer and Activator of Transcription 3 in Mice and Cynomolgus Monkeys. Nucleic Acid Ther. 2013;23(3):213-27.
159. Zinzalla G, Haque MR, Basu BP, Anderson J, Kaye SL, Haider S, et al. A novel small-molecule inhibitor of IL-6 signalling. Bioorg Med Chem Lett. 2010;20(23):7029-32.
160. Fu H. Protein - Protein Interactions: Methods and Applications (Methods in Molecular Biology): Humana Press; 2004 edition; 2004.
161. Cardullo RA. Theoretical principles and practical considerations for fluorescence resonance energy transfer microscopy. Methods Cell Biol. 81. United States2007. p. 479-94.
162. Kretzschmar AK, Dinger MC, Henze C, Brocke-Heidrich K, Horn F. Analysis of Stat3 (signal transducer and activator of transcription 3) dimerization by fluorescence resonance energy transfer in living cells. Biochem J. 377. England2004. p. 289-97.

163. Ullman EF, Schwarzberg M, Rubenstein KE. Fluorescent excitation transfer immunoassay. A general method for determination of antigens. J Biol Chem. 1976;251(14):4172-8.
164. Clegg RM. Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol. 1992;211:353-88.
165. Dechow TN, Pedranzini L, Leitch A, Leslie K, Gerald WL, Linkov I, et al. Requirement of matrix metalloproteinase-9 for the transformation of human mammary epithelial cells by Stat3-C. P Natl Acad Sci USA. 2004;101(29):10602-7.
166. Yang J, Huang J, Dasgupta M, Sears N, Miyagi M, Wang B, et al. Reversible methylation of promoter-bound STAT3 by histone-modifying enzymes. Proc Natl Acad Sci U S A. 2010;107(50):21499-504.
167. Belton A, Xian L, Huso T, Koo M, Luo LZ, Turkson J, et al. STAT3 inhibitor has potent antitumor activity in B-lineage acute lymphoblastic leukemia cells overexpressing the high mobility group A1 (HMGA1)-STAT3 pathway. Leuk Lymphoma. 2016;57(11):2681-4.
168. Liu LJ, Leung KH, Chan DS, Wang YT, Ma DL, Leung CH. Identification of a natural product-like STAT3 dimerization inhibitor by structure-based virtual screening. Cell Death Dis. 2014;5:e1293.
169. Wang P, Henning SM, Heber D. Limitations of MTT and MTS-Based Assays for Measurement of Antiproliferative Activity of Green Tea Polyphenols. PLOS ONE. 2010;5(4):e10202.
170. Wilson ML, Macnab RM. Overproduction of the MotA protein of Escherichia coli and estimation of its wild-type level. J Bacteriol. 1988;170(2):588-97.
171. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera--a visualization system for exploratory research and analysis. J Comput Chem. 2004;25(13):1605-12.
172. Hawkins GD, Cramer CJ, Truhlar DG. Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. The Journal of Physical Chemistry. 1996;100(51):19824-39.
173. Park IH, Li C. Characterization of molecular recognition of STAT3 SH2 domain inhibitors through molecular simulation. J Mol Recognit. 2011;24(2):254-65.
174. Fink AL. Protein aggregation: folding aggregates, inclusion bodies and amyloid. Fold Des. 3. England1998. p. R9-23.
175. Turkson J, Zhang SM, Palmer J, Kay H, Stanko J, Mora LB, et al. Inhibition of constitutive signal transducer and activator of transcription 3 activation by novel platinum complexes with potent antitumor activity. Molecular Cancer Therapeutics. 2004;3(12):1533-42.
176. Hsieh F-C, Cheng G, Lin J. Evaluation of potential Stat3-regulated genes in human breast cancer. Biochemical and Biophysical Research Communications. 2005;335(2):292-9.
177. Alvarez JV, Greulich H, Sellers WR, Meyerson M, Frank DA. Signal transducer and activator of transcription 3 is required for the oncogenic effects of non-small-cell lung cancer-associated mutations of the epidermal growth factor receptor. Cancer Res. 66. United States2006. p. 3162-8.

9 Appendices

9.1 HIS-STAT3β$_{TC}$ sequencing data

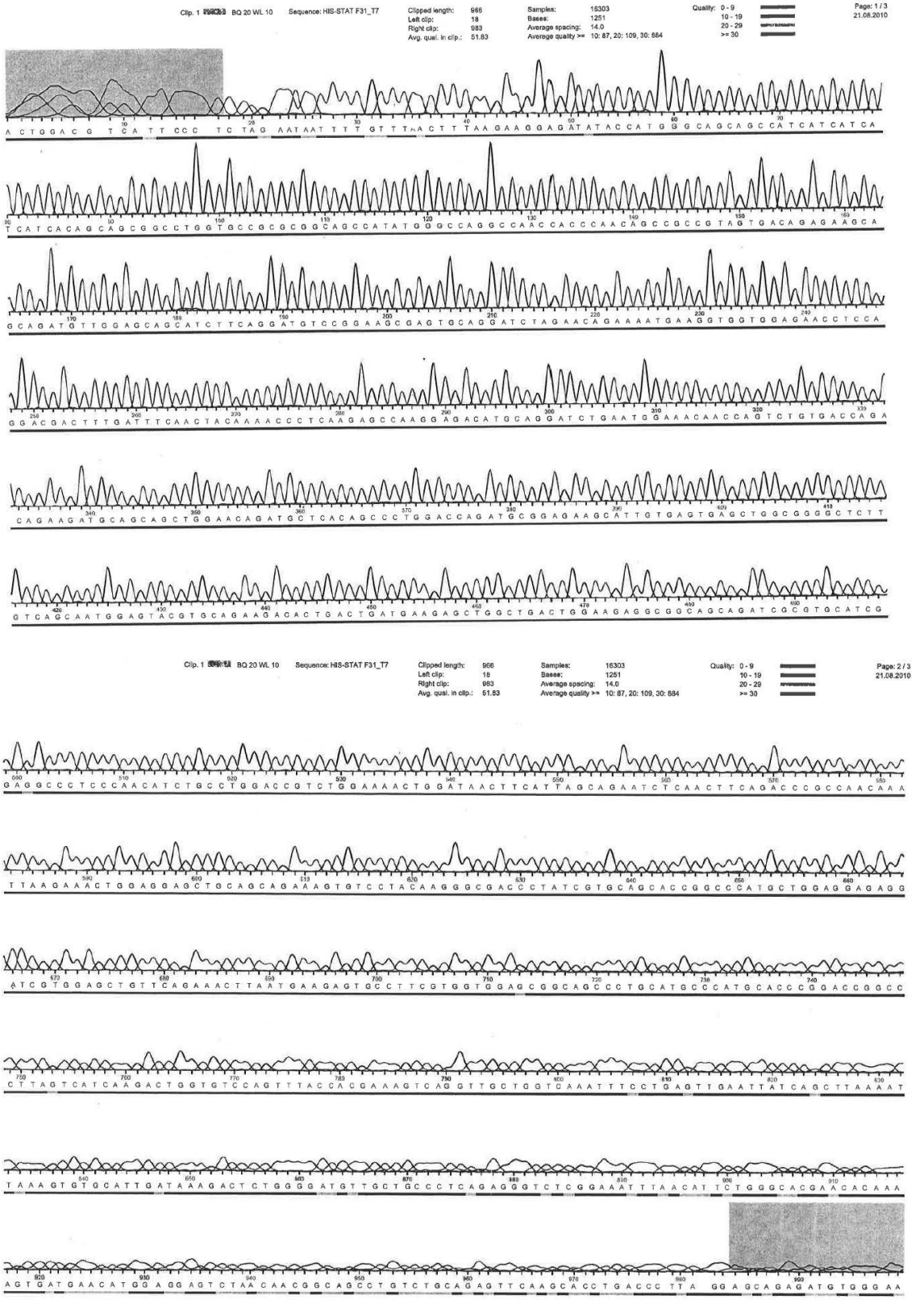
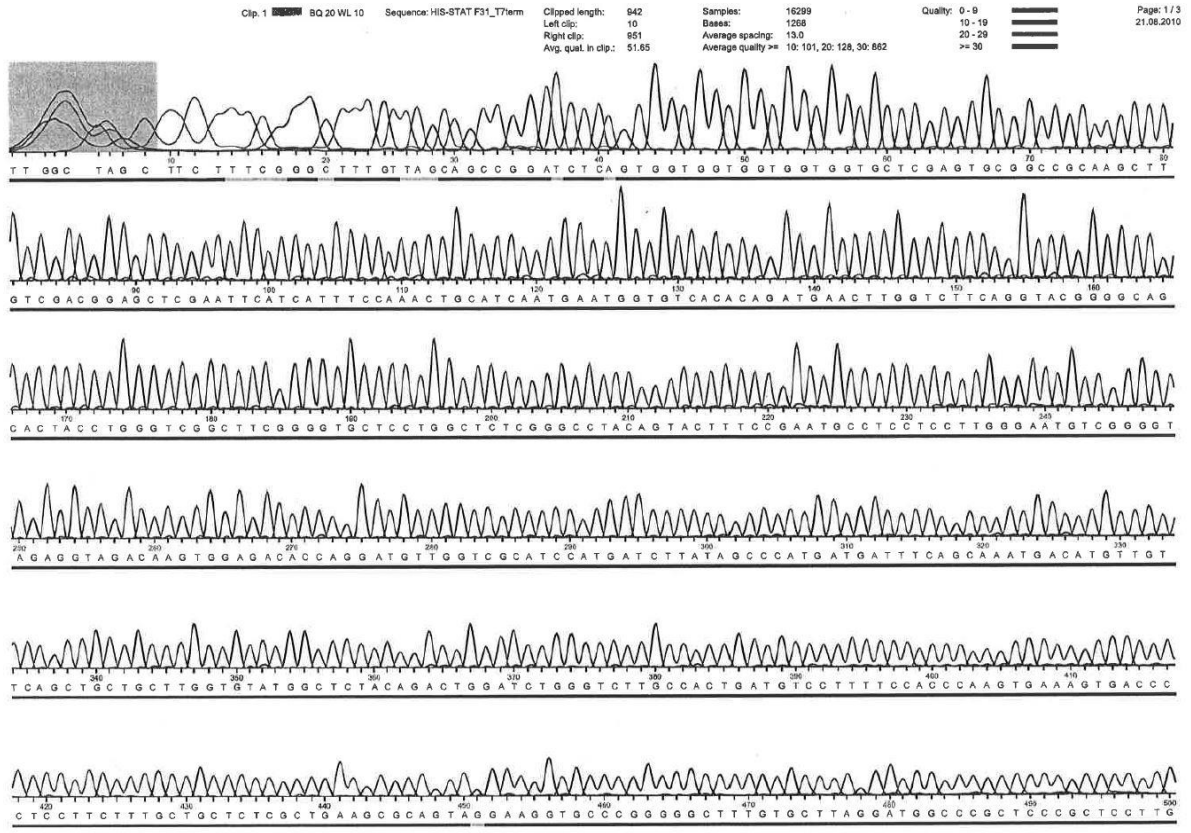


Figure 9.54 Forward T7 primer sequencing data for HIS-STAT3β_{TC} pet 28C vector



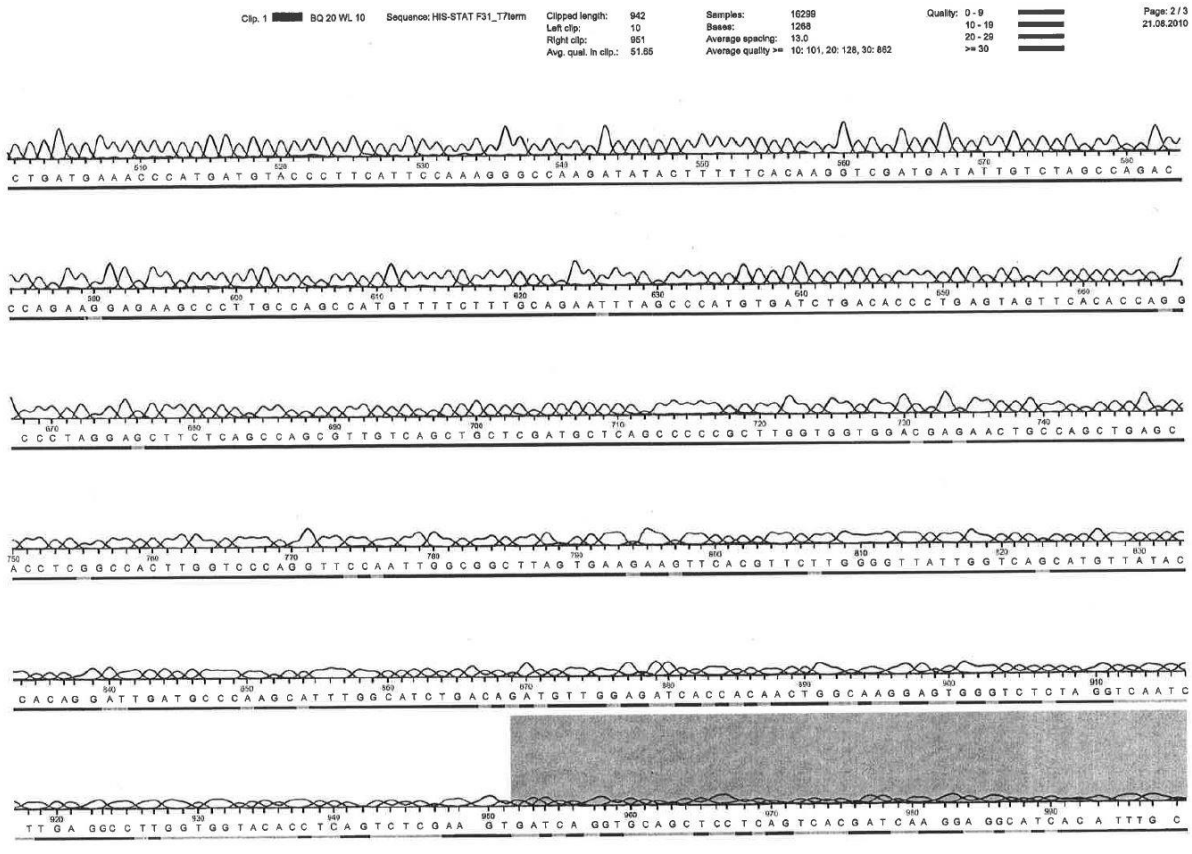


Figure 9.55 Reverse T7-term primer sequencing data for HIS-STAT3β_{TC} pet 28C vector

9.2 His-STAT3βTC amino acid sequence

M G S S HHHHHH S S G L L P R G S H M GQAN HPTAAVVTEK QQMLEQHLQD VRKRVQDLEQ KMKVVENLQD DFDFNYKTLK

SQGDMQDLNG NNQSVTRQKM QQLEQMLTAL DQMRRSIVSE LAGLLSAMEY VQKTLTDEEL

ADWKRRQQIA CIGGPPNICL DRLENWITSL AESQLQTRQQ IKKLEELQQK VSYKGDPIVQ

HRPMLEERIV ELFRNLMKSA FVVERQPCMP MHPDRPLVIK TGVQFTTKVR LLVKFPELNY

QLKIKVCIDK DSGDVAALRG SRKFNILGTN TKVMNMEESN NGSLSAEFKH LTLREQRCGN

GGRANCDASL IVTEELHLIT FETEVYHQGL KIDLETHSLP VVVISNICQM PNAWASILWY

NMLTNNPKNV NFFTKPPIGT WDQVAEVLSW QFSSTTKRGL SIEQLTTLAE KLLGPGVNYS

GCQITWAKFC KENMAGKGFS FWVWLDNIID LVKKYILALW NEGYIMGFIS KERERAILST

KPPGTFLLRF SESSKEGGVT FTWVEKDISG KTQIQSVEPY TKQQLNNMSF AEIIMGYKIM

DATNILVSPL VYLYPDIPKE EAFGKYCRPE SQEHPEADPG SAAPYLKTKF ICVTP

F I D A V W K STOP

AA = 617

Figure 9.56 Expressed HIS-STAT3β$_{TC}$ amino acid sequence

9.3 Protparam protein characterisation data (HIS-STAT3β_{TC})

Number of amino acids: 617

Molecular weight: 70406.9

Theoretical pI: 7.58

Amino acid composition: Submit

Ala (A)	30	4.9%
Arg (R)	26	4.2%
Asn (N)	31	5.0%
Asp (D)	26	4.2%
Cys (C)	11	1.8%
Gln (Q)	41	6.6%
Glu (E)	44	7.1%
Gly (G)	35	5.7%
His (H)	16	2.6%
Ile (I)	36	5.8%
Leu (L)	61	9.9%
Lys (K)	45	7.3%
Met (M)	22	3.6%
Phe (F)	23	3.7%
Pro (P)	28	4.5%
Ser (S)	40	6.5%
Thr (T)	36	5.8%
Trp (W)	12	1.9%
Tyr (Y)	15	2.4%
Val (V)	39	6.3%
Pyl (O)	0	0.0%
Sec (U)	0	0.0%
(B)	0	0.0%
(Z)	0	0.0%
(X)	0	0.0%

Total number of negatively charged residues (Asp + Glu): 70
Total number of positively charged residues (Arg + Lys): 71

Atomic composition:

Carbon	C	3133
Hydrogen	H	4954
Nitrogen	N	856
Oxygen	O	921
Sulphur	S	33

Formula: $C_{3133}H_{4954}N_{856}O_{921}S_{33}$
Total number of atoms: 9897

Extinction coefficients:

Extinction coefficients are in units of M^{-1} cm^{-1}, at 280 nm measured in water.

```
Ext. coefficient    88975
Abs 0.1% (=1 g/l)   1.264, assuming all pairs of Cys residues form cystines


Ext. coefficient    88350
Abs 0.1% (=1 g/l)   1.255, assuming all Cys residues are reduced
```

Estimated half-life:

```
The N-terminal of the sequence considered is M (Met).

The estimated half-life is: 30 hours (mammalian reticulocytes, in vitro).
                            >20 hours (yeast, in vivo).
                            >10 hours (Escherichia coli, in vivo).
```

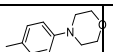
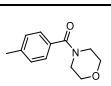
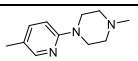
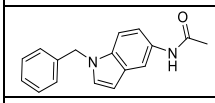
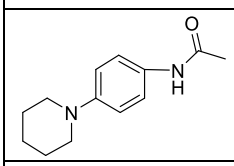
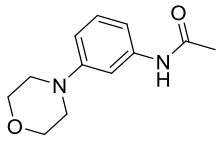
Instability index:

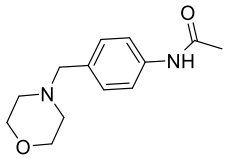
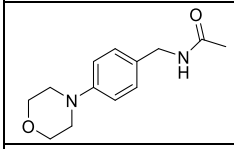
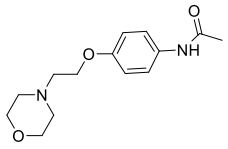
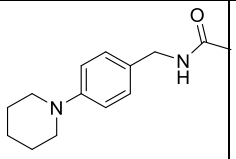
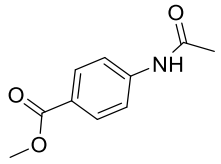
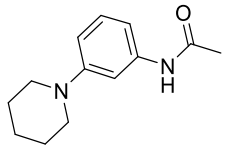
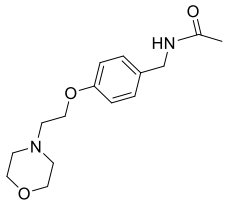
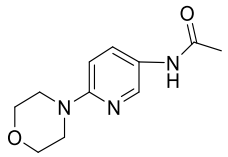
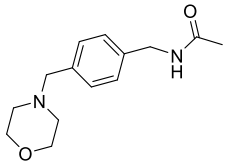
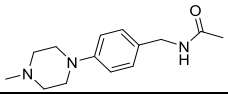
```
The instability index (II) is computed to be 42.21
This classifies the protein as unstable.
```

Aliphatic index: 84.51

Grand average of hydropathicity (GRAVY): -0.392

9.4 Compound library nomenclature and R' groups

R1 \ R2	Br										
	57-06 **1a**	57-04, SS-04 **2a**	57-05 **2b**	57-13 **2c**	AG-04	AG-07	AG-08		SS-05 AG	SS-08	SS-07
	57-07 **1b**	57-10 **2d**									
	57-09										
	57-16 **1c**										
	57-17 **1d**										

	57-18 **1e**										
	57-21 **1f**										
	57-22 **1g**	57-39, SS-06 **2e**			65-04 **2f**	65-05 **2g**	65-08 **2h**	65-09 **2i**	65-10 **2j**		SS 11
	57-23 **1h**										
	57-24 **1i**										
	57-25 **1j**										
	57-26 **1k**										
	57-27 **1l**										
	57-28 **1m**										
	57-29 **1n**										

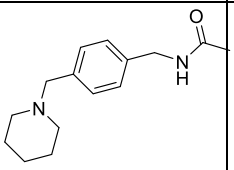
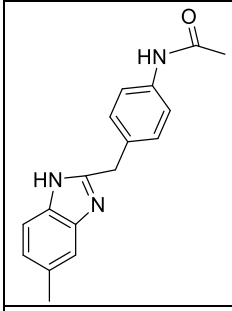
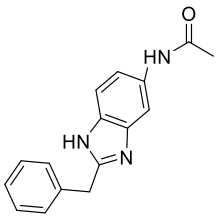
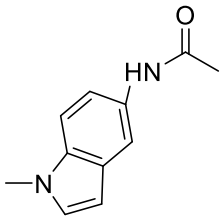
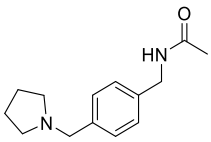
	57-30 **1o**	65-12			65-11 **2k**	65-14	65-13	65-15	SS-09	65-16	SS-10
	57-31 **1p**										
	57-32 **1q**										
	57-33 **1r**										
	57-34 **1s**										

Table 9.20 Compound nomenclature with R^1 and R^2 groups

9.5 Molecular modelling summary table

Molecule	GRID Score Protein (kcal/mol)	Interfere with Y705?	GBSA Score Protein (kcal/mol)	Interfere with Y705?	Ranking	GRID Score DNA (kcal/mol)	GBSA Score DNA (kcal/mol)	Protein (P)/DNA (D)?
KSN-57-3	-46.10	N	-51.54	N		-44.98	-51.07	P
KSN-57-4	-53.46	Y	-54.63	Y		-63.01	-70.18	D
KSN-57-5	-55.50	N	-62.65	N		-54.01	-60.81	P
KSN-57-6	-46.57	N	-51.96	N (binds to back of protein)		-48.50	-50.25	P
KSN-57-7	-58.76	Y	-27.33	Y		-61.59	-69.09	D
KSN-57-9	-58.91	N	-66.81	N		-56.28	-64.55	P
KSN-57-10	-70.05	Y	-64.42	Y	9	-58.67	-65.25	P
KSN-57-13	-53.40	N	-60.68	N		-49.44	-55.56	P
KSN-57-16	-57.07	Y	-61.52	Y	16	Failed	Failed	P
KSN-57-17	-58.56	Y	-43.80	N		Failed	Failed	P
KSN-57-18	-59.55	Y	-56.13	Y		Failed	Failed	P
KSN-57-21	-58.18	Y	-65.99	Y	13	-55.62	-17.24	P
KSN-57-22	-62.80	Y	-68.65	N		Failed	Failed	P
KSN-57-23	-67.34	Y	-70.49	Y	6	-53.68	-61.21	P
KSN-57-25	-58.18	Y	-62.50	Y	14	-31.95	-32.03	P
KSN-57-26	-69.54	Y	-73.98	Y	2	-21.59	Failed	P
KSN-57-27	-59.14	Y	-64.79	N		-49.08	-56.44	P
KSN-57-28	-61.29	Y	-59.11	Y	15	-33.89	-45.90	P
KSN-57-29	-62.15	Y	-71.22	Y	10	-37.57	-53.68	P
KSN-57-30	-65.74	Y	-63.04	Y	12	-27.31	-50.40	P
KSN-57-31	-61.16	Y	-66.73	N		Failed	-49.63	P
KSN-57-32	-66.29	N	-72.33	N		-43.35	-70.84	P

KSN-57-33	-53.96	N	-57.33	N		-61.14	-68.20	D
KSN-57-34	-59.78	N	-66.21	N		-52.55	-40.80	P
KSN-57-36	-65.41	Y	-73.26	Y	7	Failed	Failed	P
KSN-57-37	-65.55	Y	-47.09	N		-55.80	-33.50	D
KSN-57-38	-69.19	Y	-23.76	N		-63.67	-73.21	P
KSN-57-39	-70.37	Y (interferes with P and Y)	-70.47	Y (interferes with P and Y)	5	-60.16	-70.56	P
KSN-57-40	-65.44	N	-60.19	N		-55.06	-62.59	P
KSN-57-41 (AG)	-57.44	N	-64.99	N		-55.19	-63.34	P
KSN-57-42 (AG)	-59.84	N	-67.87	N		-57.94	-66.10	P
KSN-57-43 (AG)	-57.12	N	-64.63	N		-50.20	-52.69	P
KSN-57-44 (AG)	-59.94	N	-67.29	N		-55.31	-63.85	P
KSN-57-45 (AG)	-53.95	N	-59.78	N		-62.68	-63.57	D
KSN-65-01 (AG)	-57.30	N	-66.06	N		-53.77	-61.68	P
KSN-65-02 (AG)	-57.20	N	-64.41	N		-51.60	-56.67	P
KSN-65-03 (AG)	-58.72	N	-66.13	N		-60.34	-65.63	P
KSN-65-04	-62.78	Y (interferes with P and Y)	-68.27	Y (interferes with P and Y)	11	-71.10	-79.50	D
KSN-65-05	-70.77	Y (interferes with P and Y)	-76.39	Y (interferes with P and Y)	1	-71.83	-80.30	P
KSN-65-06	-52.47	N	-59.44	N		-47.22	-53.06	P
KSN-65-07	-59.77	N	-68.51	N		-51.87	-58.56	P
KSN-65-08	-71.78	Y (interferes with P and Y)	-41.71	Y		-59.72	-64.97	P
KSN-65-09	-70.26	Y (interfere	-70.57	Y (interfere	4	-61.59	-13.66	P

		s with P and Y)		s with P and Y)				
KSN-65-10	-71.43	Y	-70.72	Y	3	-68.89	-67.42	P
KSN-65-11	-66.99	Y (interferes with P and Y)	-67.14	Y (interferes with P and Y)	8	-59.60	-67.45	P
SS-04	-52.50	N	-60.13	N		-58.20 (N)		
SS-05	-54.34	N	-56.41	N		-51.05 (N)		
SS-06 (57-39)	-76.11	Y (interferes with P and Y)	-74.61	Y (interferes with P and Y)		-72.99 (Y)	-80.93 (N)	
SS-07	-54.34	N	-52.51	N		-46.42 (N)		
SS-08	-53.88	N	-64.05	N		-52.21 (N)		
SS-09	-65.81	Y (interferes with P and Y)	-9.80	Y (interferes with P and Y)		-74.65 (Y)	-36.88 (Y)	
SS-10	-62.52	Y (interferes with P and Y)	-23.84	Y (interferes with P and Y)		-63.42 (N)	-59.33 (N)	
SS-11	-69.86	Y (interferes with P and Y)	-68.08	Y (interferes with P and Y)		-56.48 (N)	-67.25 (N)	
SS-12	-71.29	Y (interferes with P and Y)	+19.74	Repulsive contacts		-64.20 (N)	-75.11 (N)	

Table 9.21 Docking scores of KSN-57- and KSN-65- series of molecules (both GRID and GBSA scoring in kcal/mol) in both the SH2 and DNA binding domains.